18 July 2008 | $10

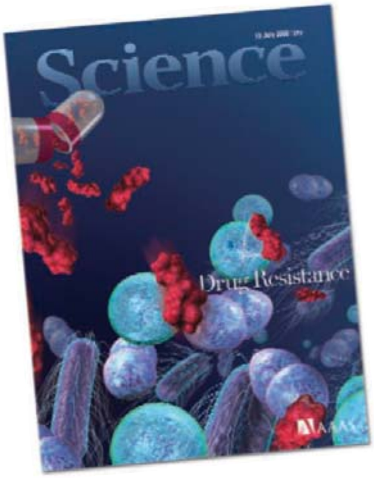
Science

Drug Resistance

AAAS

CONTENTS

Volume 321, Issue 5887



COVER

An artist's conception of the antibiotic penicillin and some of the bacteria that have developed resistance to various antibiotics. A special section beginning on page 355 explores the rise and spread of so-called bad bugs and possible interventions.

Illustration: Chris Bickel/**Science**

DEPARTMENTS

EDITORIAL

SPECIAL SECTION

Drug Resistance

INTRODUCTION

NEWS

PERSPECTIVES

>> For related online content, see the Science *Podcast, p. 311*

355

330

NEWS OF THE WEEK

NEWS FOCUS

CONTENTS continued **>>**

SCIENCE EXPRESS

www.sciencexpress.org

PLANT SCIENCE

Plant Immunity Requires Conformational Changes of NPR1 via S-Nitrosylation and Thioredoxins

Y. Tada et al.

After a pathogen invades a plant, a protein, usually kept in a multimeric state by S-nitrosylation, is dissociated by thioredoxin, freeing the monomers for defense responses.

10.1126/science.1156970

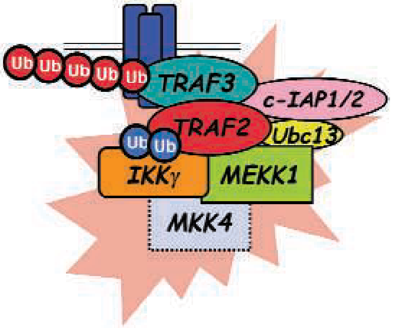
GEOCHEMISTRY

Ferruginous Conditions Dominated Later Neoproterozoic Deep-Water Chemistry

D. E. Canfield et al.

Low sulfur input caused the deeper ocean to become anoxic and rich in ferrous iron 750 million years ago, a reversal from the more oxidizing conditions of the previous 1 billion years.

10.1126/science.1154499



CELL BIOLOGY

Essential Cytoplasmic Translocation of a Cytokine Receptor–Assembled Signaling Complex

A. Matsuzawa et al.

Degradation of one member of a protein complex that forms when a cytokine receptor is activated causes the complex to move to the cytoplasm, triggering the downstream pathway.

10.1126/science.1157340

LETTERS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

347 &
417

TECHNICAL COMMENT ABSTRACTS

MICROBIOLOGY

BREVIA

OCEAN SCIENCE

Ice Scour Disturbance in Antarctic Waters 371

D. A. Smale et al.

Icebergs have increasingly scoured the coastlines along the West Antarctic Peninsula as its ice shelves and glaciers have waned, affecting benthic marine communities.

RESEARCH ARTICLE

NEUROSCIENCE

Identification of SLEEPLESS, a Sleep-Promoting Factor 372

K. Koh et al.

A search for genetic modulators of sleep in *Drosophila* identified a gene encoding a brain protein that is likely secreted and is required for recovery from sleep deprivation. *>> News story p. 334*

REPORTS

ASTRONOMY

Properties of Gamma-Ray Burst Progenitor Stars 376

P. Kumar, R. Narayan, J. L. Johnson

Analysis of the x-ray afterglow of intense gamma-ray bursts shows that the bursts result from consumption of the outer part of a dense star and define the star's rotation rate.

CONTENTS continued >>

Published by AAAS

REPORTS *CONTINUED...*

APPLIED PHYSICS

High-Resolution Scanning X-ray Diffraction Microscopy 379

P. Thibault et al.

Analysis of differences in diffraction patterns at each point along an x-ray scan of a material allows imaging of a buried structure with a resolution of 50 nanometers. *>> Perspective p. 352*

MATERIALS SCIENCE

Observations of Intergranular Stress Corrosion Cracking in a Grain-Mapped Polycrystal 382

A. King et al.

Tomographic imaging reveals that some grain boundaries in stainless steel are resistant to stress corrosion cracking, which leads to sudden brittle failure. *>> Perspective p. 349*

MATERIALS SCIENCE

Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene 385

C. Lee, X. Wei, J. W. Kysar, J. Hone

Measurements of the elastic properties of graphene agree with calculations for a defect-free material and show that it is indeed stronger than other materials.

CHEMISTRY

Measurement of the Distribution of Site Enhancements in Surface-Enhanced Raman Scattering 388

Y. Fang, N.-H. Seong, D. D. Dlott

The distribution of electric field–enhancing sites on a nanostructured substrate is measured by using the enhanced field to damage those sites.

CLIMATE CHANGE

Patagonian Glacier Response During the Late Glacial–Holocene Transition 392

R. P. Ackert Jr. et al.

Dating of a glacial moraine in southern Patagonia implies that increased precipitation caused glacier growth after a period of Northern Hemisphere cooling 11,000 years ago. *>> Perspective p. 348*

SOCIOLOGY

Electronic Publication and the Narrowing of Science and Scholarship 395

J. A. Evans

As journals become available electronically, scientists and scholars have more articles at their fingertips but cite relatively fewer, and these tend to be more recent. *>> News story p. 329*

EVOLUTION

The Evolution and Distribution of Species Body Size 399

A. Clauset and D. H. Erwin

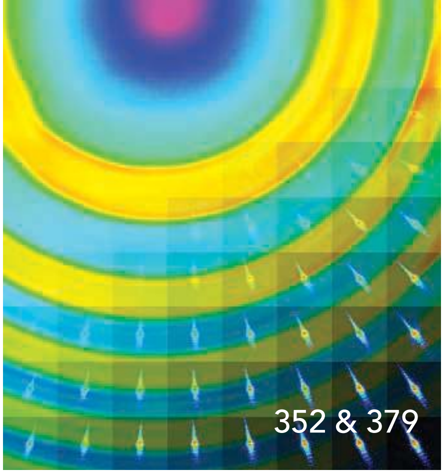
A model of evolutionary body-size changes that accounts for physical constraints and extinction risk reproduces the size distribution of land mammals from the Quaternary.

CELL BIOLOGY

Four-jointed Is a Golgi Kinase That Phosphorylates a Subset of Cadherin Domains 401

H. O. Ishikawa et al.

A newly described type of protein kinase found in the Golgi phosphorylates signaling proteins on amino acids that are destined to be within extracellular domains.



CELL BIOLOGY

Signal-Mediated Dynamic Retention of Glycosyltransferases in the Golgi 404

L. Tu, W. C. S. Tai, L. Chen, D. K. Banfield

Glycosyltransferase enzymes stay in the Golgi in the face of continuing membrane traffic because a receptor links their cytoplasmic tails to a recycling coated vesicle.

IMMUNOLOGY

Anomalous Type 17 Response to Viral Infection by CD8$^+$ T Cells Lacking T-bet and Eomesodermin 408

A. M. Intlekofer et al.

Two transcription factors cooperate to ensure the correct functioning of CD8$^+$ T cells during the response to infection.

CELL SIGNALING

Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP 411

N. Sudarsan et al.

The bacterial second messenger cyclic di–guanosine monophosphate controls a wide variety of cellular functions by acting on a riboswitch motif in numerous messenger RNAs.

NEUROSCIENCE

Bottom-Up Dependent Gating of Frontal Signals in Early Visual Cortex 414

L. B. Ekstrom et al.

Higher brain centers can modulate activity in the cortical regions that directly receive visual input, but only when a visual stimulus is present.

NEUROSCIENCE

Evolutionary Origins for Social Vocalization in a Vertebrate Hindbrain–Spinal Compartment 417

A. H. Bass, E. H. Gilland, R. Baker

The conserved neural circuitry for vocal communication in fish and other tetrapods suggests that this function may have originated prior to the evolution of bony vertebrates. *>> Perspective p. 347*

NEUROSCIENCE

Orbitofrontal Dysfunction in Patients with Obsessive-Compulsive Disorder and Their Unaffected Relatives 421

S. R. Chamberlain et al.

The abnormally low activation in the frontal cortex of individuals with obsessive compulsive disorder and their close relatives may confer a risk for the disease.

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2008 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $144 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $770; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

CONTENTS continued **>>**

Science ONLINE

www.sciencemag.org

Devil destroyer.

SCIENCE**NOW**

www.sciencenow.org

HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

Tough Times for the Taz

Researchers debate the evolutionary impact of a deadly cancer in Tasmanian devils.

Answer to Carbon Emissions May Lie Under the Sea

Researchers propose injecting greenhouse gas near volcanic rock on the ocean bottom.

"Baby Boom" in a Stellar Nursery

Astronomers discover an ancient galactic star factory on overdrive.

Listeria, a human pathogen.

SCIENCE **SIGNALING**

www.sciencesignaling.org

THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

REVIEW: Host-Directed Drug Targeting of Factors Hijacked by Pathogens

A. Schwegmann and F. Brombacher

A new drug discovery paradigm focuses on identifying and targeting cellular elements of the host that are exploited by pathogens.

GLOSSARY

Find out what DILP, HRE, and OGT mean in the world of cell signaling.

EVENTS

Check out the more than 50 cell signaling–related meetings happening in the second half of 2008.

Engineering to minimize earthquake damage.

SCIENCE **CAREERS**

www.sciencecareers.org/career_development

FREE CAREER RESOURCES FOR SCIENTISTS

Do You Wanna Be a VAP?

L. Malisheski

A visiting assistant professor position can be a step forward on some career paths.

Tooling Up: Review Your Career

D. Jensen

It may be time to shake things up with a career review.

Designing for the Next Quake

A. Saini

Earthquake engineers study how to avoid seismic destruction.

Science Careers Seeks Bloggers

J. Austin

We are looking for a few people with interesting things to say about their careers in science.

SCIENCE**PODCAST**

www.sciencemag.org/multimedia/podcast

FREE WEEKLY SHOW

Download the 18 July *Science* Podcast to hear about pathogenic fungi, resistance genes in natural environments, extremely multidrug-resistant bacteria, and more.

>> Drug Resistance section p. 355

Separate individual or institutional subscriptions to these products may be required for full-text access.

CREDITS: (*SCIENCE NOW*) MENNA JONES; (*SCIENCE CAREERS*) NSF/WILLIAM A. MITCHELL; (*SCIENCE SIGNALING*) PHOTOTAKE, INC./ALAMY

EDITED BY STELLA HURTLEY

<< Talking Fish

Although speech seems a particularly human characteristic, vocalizations that impart social and environmental information are common to a variety of other animals, including birds, frogs, and even fish. **Bass *et al.*** (p. 417; see the Perspective by **Margoliash and Hale**) studied the development of larval Batrachoidid fish, the adults of which use a complex pattern of vocalizations. Analysis of the developing hindbrain, particularly the eighth rhombomere, showed the beginnings of the vocal motor nucleus. The development of this vocal pacemaker circuit in fish reflects similar patterns of development known from other vertebrates. Thus, the brain circuitry driving vocalizations may have its origins far back in the evolution leading to bony vertebrates.

Sleepless in *Drosophila*

All animals sleep, and the more they are awake, the stronger the drive to sleep. To better understand the process of sleep, **Koh *et al.*** (p. 372; see the news story by **Youngsteadt**) screened mutagenized *Drosophila* for genes involved in sleep regulation. They found one—*sleepless*—that is required for normal sleep; without *sleepless*, flies sleep much less, about 20% of normal. *Sleepless* is also required for rebound sleep after prolonged waking. *Sleepless* is an allele of *quiver*, a gene that modulates the K$^+$-channel activity encoded by *Shaker*, which also affects sleep. K$^+$ channel–associated neuronal sensitivity may thus play a role in the control of sleep, and the SLEEPLESS protein may signal the drive to sleep by decreasing membrane excitability.

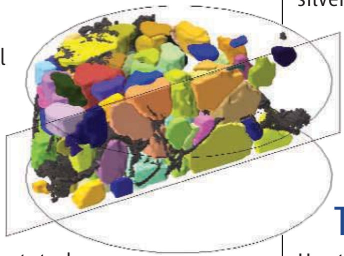
Afterglow

Gamma-ray bursts, the most energetic emissions in the universe, are thought to be produced when a black hole consumes a rapidly rotating, high-mass star. Many bursts are followed by an extended x-ray afterglow. **Kumar *et al.*** (p. 376, published online 26 June) analyze this afterglow and, assuming that it represents continued emissions from the stars, use it to determine the properties of the consumed stars. The analysis of three stars characterizes their rotation speeds and shows that only a few solar masses are consumed in the outburst, even though the stars may have been several times as large.

Striving for Perfection

The mechanical properties of a material rarely achieve theoretical or ideal properties, due to defects that are formed during synthesis or processing. For materials with nanoscale dimensions, the gap between the ideal and real cases can close, because it is easier to make materials that are close to being defect-free. **Lee *et al.*** (p. 385) measure the elastic properties and breaking strength of graphene membranes, which are one-atom-thick carbon sheets, and find values that agree with recent simulations and calculations. The material can be deformed well beyond the linear regime, and graphene is one of the strongest materials ever measured.

Not Cracking Under the Strain



Ductile materials will deform before failure, while brittle ones show sudden cracking and rupture and thus give little warning of the impending doom. Most steels will show ductile failure, but if placed under a constant tensile stress and exposed to the right (or wrong) chemical environment, they can suddenly undergo brittle failure through the formation of stress corrosion cracking. **King *et al.*** (p. 382; see the Perspective by **Stierle**) use diffraction contrast tomography to track intergranular stress corrosion cracking. Special grain boundaries could be observed that are resistant to the cracking process due to the formation of bridging ligaments that retain some ductility within the material.

Details from Damage

Surface-enhanced Raman scattering (SERS) is a well-established phenomenon whereby incident laser fields are locally enhanced a millionfold or more by metallic substrates sharply structured at the nanoscale. The effect has been used for sensitive molecular sensing applications, but approaches toward optimizing substrate geometries for maximal enhancement remain somewhat empirical. A particular challenge has been quantifying the distribution of enhancement magnitudes across multiple sites on a given surface. **Fang *et al.*** (p. 388, published online 26 June) explore this distribution by using the field enhancement to induce damage (presumably by ionization) of adsorbed molecules on a widely studied SERS substrate of silver-coated nanoparticles. By steadily ramping up the energy of an incident laser pulse, they progressively damage molecules at sites with diminishing enhancement factors, observing the depletion with a low-intensity probe pulse.

The Invisible Past

Has the move to electronic publication changed authors' styles of searching and citing the literature? **Evans** (p. 395) reports that researchers are referencing more narrowly than in the past, citing fewer, more recent references. A database of 34 million articles from journals that became available online between 1998 and 2005 was analyzed for the number of articles (from a given journal) cited by any other articles in a given publication year. The results were consistent over time and were not journal- or subfield-specific. Perhaps because of the lack of hyperlinks and efficient electronic indexing, individuals searching through the print literature may be exposed to a broader set of references and ideas.

Continued on page 315

CREDITS (TOP TO BOTTOM): HERB SEGARS/ANIMALS ANIMALS; KING ET AL.

Continued from page 313

Phosphorylation Inside-Out

A signaling pathway called the Hippo pathway has important roles in the control of tissue organization in development and in the regulation of organ size in the fruit fly. Hippo is a protein kinase that appears to pass signals from atypical cadherin molecules Fat and Dachsous at the cell surface to another intracellular kinase, Warts, which then regulates a transcriptional coactivator. Another protein somewhat distantly related to the protein kinases, Four-jointed, has also been implicated genetically in the pathway. **Ishikawa *et al.*** (p. 401) provide evidence that Four-jointed, which is localized to the Golgi complex, is indeed a kinase and that it appears to phosphorylate the cadherin domains of Fat and Dachsous, which will become extracellular domains at the cell surface.

Doubled-Up Decision

CD8$^+$ T cells are major contributors to cell-mediated immunity to virally infected cells and tumors. Like their helper CD4$^+$ counterparts, these cells rely on the transcriptional regulator T-bet for their correct development. Recently, a second factor called eomesodermin has also been found to control CD8 functions. **Intlekofer *et al.*** (p. 408) now find that without both factors, CD8$^+$ T cells fail to develop their normal cell-mediated functions and instead secrete the inflammatory cytokine IL-17, which has been recently characterized in helper T cells. This secretion of IL-17 caused significant pathology in a mouse model of viral infection, suggesting that both transcription factors play a crucial role in maintaining appropriate cell-mediated responses to infection.

Absence of Cooling

The Younger Dryas was an approximately 1300-year period that interrupted the warming of the last deglaciation, during which markedly colder conditions clearly recurred in many parts of the Northern Hemisphere. Whether or not the Southern Hemisphere experienced concurrent cooling is an open question. **Ackert *et al.*** (p. 392; see the Perspective by **Lowell and Kelly**) measured cosmic-ray exposure ages of a glacial moraine in southern Patagonia in order to determine whether or not the glacial advance that created it occurred during the Younger Dryas chronozone. The moraine was deposited soon after the end of the Younger Dryas, and the glacier grew in response to more precipitation, not because of regional cooling. This suggests that temperatures in the Southern Hemisphere did not drop like those in the North during the Younger Dryas.

Simultaneous Brain Imaging and Microstimulation

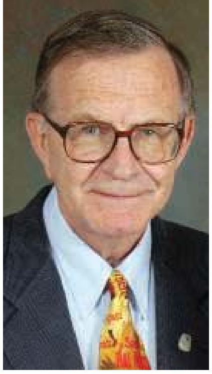
Until now, functional brain-imaging studies have focused on how regions are activated by a particular stimulus or cognitive task. However, how nodes within a functional network causally interact with each other is still poorly understood. **Ekstrom *et al.*** (p. 414) used a novel combination of chronic intracortical microstimulation and functional magnetic resonance imaging in awake, behaving monkeys to study the impact of frontal top-down signals on incoming sensory information. Frontal eye fields could modulate early visual areas only in the presence of a visual stimulus, whereas higher-order visual areas could be modulated independent of visual stimulation.

Orbitofrontal Obsessions

Obsessive-compulsive disorder is a debilitating neuropsychiatric condition characterized by recurrent intrusive thoughts (obsessions) and repetitive rituals (compulsions) often performed according to rigid rules. Abnormal function of the orbitofrontal cortex is central to neurobiological models of this disease. However, it is unclear whether these abnormalities are due to the symptoms of the disorder or represent a vulnerability marker also existing in people at increased genetic risk. In a well-validated brain-imaging study, **Chamberlain *et al.*** (p. 421) observed reduced activation of the orbitofrontal cortex during a reversal learning task in patients with obsessive-compulsive disorder and their unaffected first-degree relatives compared to normal controls. This deficit in activation may thus represent an endogenous predisposing factor for obsessive-compulsive disorder.

CREDIT: MOMATIUK-EASTCOTT/CORBIS

Published by AAAS



Glenn Schweitzer is the director of Eurasian Programs at the U.S. National Academies in Washington, DC. E-mail: gschweitzer@nas.edu

Engaging Russian Scientists

RUSSIA NO LONGER NEEDS ASSISTANCE FROM THE WEST TO SHORE UP ITS SCIENCE AND TECHnology (S&T) base. Its gross domestic product is $1.4 trillion and increasing at an annual rate of almost 9%. Investment in nanotechnology is on track to reach $6 billion during the next several years. The research budget of the Russian Academy of Sciences is six times larger than in 2001, and research funds are on the rise throughout the ministries.

But the United States, and indeed the entire world, needs Russian assistance to address global challenges—to expand energy supplies and promote energy-efficient technologies, to protect public health and the environment, and to prevent nuclear proliferation and terrorism. International partnerships can build on successes of the past, benefiting all participants. Also, engagement promotes transparency, while encouraging Russia to be a central S&T player for achieving common global goals.

Unfortunately, the U.S. government is still mired in the outmoded concept of foreign assistance as the basis for relations with Russia. During my visits in June 2008 to the Institute of Catalysis and the Institute of Nuclear Physics in Novosibirsk, directors and researchers bemoaned the atrophy of linkages with U.S. scientists. For them, money is not the primary motivation for cooperation, because they have well-endowed clients in Russia, China, and Europe. They simply want to work at the forefront of technology with U.S. counterparts. Subsequent visits to other leading institutes in Moscow that deal with epidemiology, nuclear contamination, and geological mapping underscored the growing Russian view that U.S. colleagues are losing interest just as Russian capabilities are growing.

The U.S. government still supports efforts to reduce Russia's nuclear arsenal and contain nuclear materials in secure locations, with the Department of Energy's (DOE's) commitment of about $600 million for 2008. The National Aeronautics and Space Administration continues its partnership with Russia to support the international space station and related activities. The U.S. Agency for International Development provides modest support to combat HIV/AIDS. And in the private sector, U.S. investment in Russia is increasing, although hardly at a level commensurate with market and technological opportunities.

Aside from these bright spots, the level of U.S. support for bilateral cooperation is not encouraging. In February, the Department of Commerce closed its Business Information Services for the Newly Independent States that had facilitated transactions of about $4.5 billion over 16 years, including many investments in technology-related activities. In March, the National Science Foundation terminated nearly 50 years of support for the National Academy of Sciences' scientist-exchange program with Russia and other states in the region. The agreement to expand civil nuclear power cooperation that was signed in May is in trouble in the U.S. Congress. The Civilian Research and Development Foundation has reduced its funding for Russia, although it has succeeded in encouraging increased Russian contributions to projects. The Departments of State and Defense are reducing support for biology-related nonproliferation activities as they increase programs in other countries. And the DOE's long-standing research cooperation with the Russian Academy of Sciences is almost dormant. Very disheartening is the limited U.S. effort to launch projects pursuant to the U.S.-Russian bilateral S&T agreement. The few current projects hardly represent a credible degree of cooperation between two leaders in S&T.

U.S. agencies and scientists often cite lack of funds as a reason for reduced scientific exchanges with Russia. Although this is true, an underlying cause is the failure to recognize how Russian science can become a more positive force on the world scene. This needs to change. One solution is to expand efforts under the S&T agreement in areas such as nanotechnology and biomedical science. Another step is to firmly embed scientific cooperation in deliberations of the G8 nations, rather than raise the issue on an ad hoc basis, as has been done so often in the past. Engagement with Russia will hopefully be a more prominent issue at the 2009 G8 Summit in Italy. Times change, but cooperation remains important.

– Glenn Schweitzer

10.1126/science.1162063

CREDIT: GETTY IMAGES

EDITORS' CHOICE

EDITED BY GILBERT CHIN AND JAKE YESTON

CHEMISTRY

Peering down the Drain

Recent observations that excreted estrogenic compounds derived from pharmaceuticals can harm fish at concentrations in the ng/liter range have driven researchers with increasing urgency to track the path of these microcontaminants down the drain. Johnson *et al.* review the various pros and cons of analytical sampling versus modeling approaches toward understanding precisely what flows from the sewer into the wider world. Sampling might seem the most accurate option, but in practice, field conditions vary widely over time and space, necessitating multiple withdrawals; compounds may degrade between acquisition and analysis; and techniques may lack the requisite detection sensitivity. Modeling is a daunting alternative, in light of the numerous factors that must be considered, ranging from human drug consumption and excretion trends, to variations in the effectiveness of sewage treatment protocols, to the range of hydrological features affecting flow dynamics. On the flip side, though, the authors note that models can more easily be scaled to treat diminishingly small concentrations without running into detection thresholds. In cases where physical measurements and model results could be compared, they agreed reassuringly well (often within a factor of 3 or 4), supporting the case for an integrated approach that balances the strengths of each complementary technique. — JSY

Environ. Sci. Technol. **42**, 10.1021/es703091r (2008).

MOLECULAR BIOLOGY

Trapped in an Eddy Upstream

Almost all eukaryotic genes initiate the translation of their messenger RNA (mRNA) into protein at an AUG start codon (which codes for the amino acid methionine). Ribosomes, the protein-synthesizing engines of the cell, scan from the 5′ end of an mRNA until they find the first AUG, and then start translation. Some mRNAs contain a supernumerary AUG (and associated short coding region) upstream and independent of the main AUG/coding region, and such upstream open reading frames (uORFs) have the potential to regulate the translation of the downstream gene.

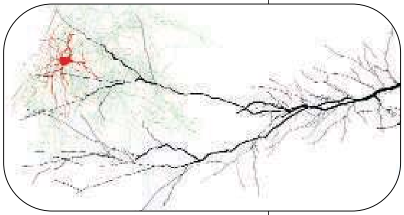
Ivanov *et al.* have found a series of conserved short uORFs associated with genes involved in polyamine synthesis, with the curious feature that they often start with a noncanonical AUU codon and hence have been overlooked in bioinformatic scans. The presence of AUU seems to be critical for the uORF to direct polyamine-directed repression of the downstream coding region; polyamines (such as spermidine) reduce the fidelity of the translation initiation complex for AUG, thus allowing increased production of the AUU uORF at the expense of the downstream polyamine biosynthesis gene—creating an autoregulatory feedback loop. The patchy distribution of these AUU uORFs across the eukaryotic phylogenetic tree suggests that they may have arisen independently on several occasions. — GR

Proc. Natl. Acad. Sci. U.S.A. **105**, 10.1073/pnas.0801590105 (2008).

NEUROSCIENCE

Fine-Tuning of Spike Timing

Neurons in layer III of the entorhinal cortex send projections along the perforant pathway that reaches area CA1 of the hippocampus. In addition to well-documented excitatory connections, there is also an important feedforward inhibitory circuit; monosynaptically activated interneurons form inhibitory synapses on CA1 pyramidal cells and thus control the timing of spiking of their target neurons. Feedforward inhibition limits the temporal summation of excitatory potentials and generates a narrow temporal window of excitability during which postsynaptic targets can fire action potentials. One important component of this feedforward inhibitory circuit is the neurogliaform cells, which frequently target the distal dendrites of excitatory neurons. Neurogliaform cells are known to be interconnected extensively through gap junctions, which has led to the hypothesis that feedforward inhibition of CA1 pyramidal cells might be highly synchronized. Price *et al.* found that stimulation of neurogliaform cells evoked GABA$_A$ receptor–mediated inhibitory postsynaptic currents (IPSCs) with a slow decay in pyramidal cells. The IPSCs also had a small but robust GABA$_B$ receptor component. Furthermore, these synapses were also subject to presynaptic GABA$_B$ receptor–mediated control. It thus makes physiological sense that these inhibitory neurogliaform-to-pyramidal cell synapses are finely tuned to control the integration time for one of the major excitatory pathways into the hippocampus. — PRS



Neurogliaform cell (red-green) and postsynaptic pyramidal cell (black).

J. Neurosci. **28**, 6974 (2008).

CREDITS (TOP TO BOTTOM): KLAUS LEIDORF/ZEFA/CORBIS; PRICE *ET AL.*, *J. NEUROSCI.* **28**, 6974 (2008)

Published by AAAS

GENETICS

Adaptive Differentiation in Barley

Variation in response to local conditions that affect growth and reproduction is a crucial means by which plants optimize their fitness. However, the underlying genetic loci that control differences between populations are generally unknown. Verhoeven *et al.* have investigated the local adaptation in two wild barley populations by measuring the response to habitat in lines with mapped quantitative trait loci (QTLs) affecting flowering time, relative growth rate, and seed weight. When individuals were transplanted reciprocally between environments, there were differences in the degree of selection on QTLs affecting flowering time, suggesting that it is a target of habitat-specific natural selection and that this adaptation may contribute to population-level divergence. — LMZ

Mol. Ecol. **17**, 3416 (2008).

MATERIALS SCIENCE

A Microfluidic Construction Kit

The field of microfluidics has blossomed as chemists and engineers have devised clever ways to handle small fluid volumes. Although many approaches exist for making devices, they often include lithographic or printing techniques. To overcome this limitation, Rhee and Burns show the feasibility of a microfluidic construction kit where individual grids are assembled by hand on a substrate. The grids range in size from 4 to 8 mm^2 and include units for fluid inlet and outlet, channels for mixing or separation, small or large chambers for sample collection, and valves and culture beds for growing cells. Grids can be placed on bare glass or on a surface coated with a thin polymer layer to improve adhesion. For better bonding, curing agents can be used to fuse the grids to the substrate and each other. Notched or covered grids, though somewhat more complex, can be used to improve grid alignment. The authors envision that these kits can expand the use of microfluidics by non-experts, particularly in the biological sciences. — MSL

Lab on a Chip **8**, 10.1039/b805137b (2008).

CHEMISTRY

Gilding the Superatom Model

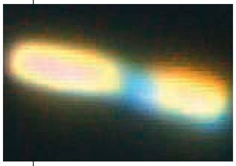
When metal atoms bind together in the gas phase to form clusters, they tend to gather preferentially in certain discrete numbers. This tendency has been rationalized with a superatom model, in which the addition of atoms to a cluster is analogous to the filling of valence electron shells in atoms themselves, so that when a critical number of atoms donate electrons to the cluster, they effectively form the equivalent of a noble gas configuration. Extending the model to clusters crystallized in solution seemed more complicated, given the requirement in that context of coordinating ligands for steric protection. However, Walter *et al.* show, using density functional theory, that the superatom model straightforwardly accounts for the particular stability of two crystallographically characterized gold clusters, the thiolate-coordinated Au_{102} and phosphine/halide-coordinated Au_{39}, as well as several smaller examples. The theory also predicts stability of 44- and 75-membered clusters that are yet to be fully structurally characterized. — JSY

Proc. Natl. Acad. Sci. U.S.A. **105**, 9157 (2008).

BIOCHEMISTRY

Mobile Intron Meets Magic Spot

Early on, introns were usually thought of as useless pieces of transcribed RNA that needed to be excised before the product RNA could become fully active. The later identification of intron-encoded species, such as microRNAs, has emphasized the utility of anything and everything that a cell carries around. In *Lactococcus lactis*, there is a mobile group II intron that consists of the catalytic LtrB RNA and the intron-encoded protein LtrA. The protein serves to stabilize the active conformation of the RNA, which splices itself out of transcripts, and also supplies a reverse transcriptase activity, which enables LtrB to insert itself at vulnerable sites in genomic DNA. One such site is the origin of replication locus (*oriC*) that in *Escherichia coli* is located at the ends of each rod-shaped cell.

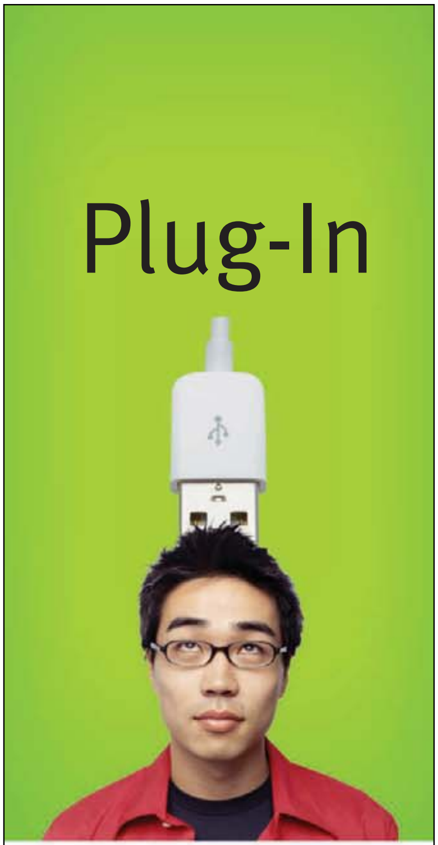


Polyphosphate (orange) in *E. coli*.

Zhao *et al.* show that LtrA localizes to the poles as well and thus accounts for the preferential integration of LtrB at *oriC*. They also find, surprisingly, that LtrA binds to polyphosphate, a curious metabolite that increases under conditions of stress and is degraded by the enzyme that synthesizes ppGpp (magic spot); this interaction has the consequence of spreading LtrA and other nucleic acid–binding proteins throughout the cell when nutrients become scarce. Whether this diffusion of polar components is the long-sought function of polyphosphate remains to be determined. — GJC

PLoS Biol. **6**, e150 (2008).

CREDIT: ZHAO ET AL., PLOS BIOL. 6, E150 (2008)



www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

Subscription Services For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store http://www.apisource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/eppstore/aaas; Bank of America MasterCard 1-800-833-6262 priority code FAA3YU; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services http://www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

science_editors@aaas.org	(for general editorial queries)
science_letters@aaas.org	(for queries about letters)
science_reviews@aaas.org	(for returning manuscript reviews)
science_bookrevs@aaas.org	(for book review queries)

Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science, engineering, and innovation throughout the world for the benefit of all people. The goals of the association are to: enhance communication among scientists, engineers, and the public; promote and defend the integrity of science and its use; strengthen support for the science and technology enterprise; provide a voice for science on societal issues; promote the responsible use of science in public policy; strengthen and diversify the science and technology workforce; foster education in science and technology for everyone; increase public engagement with science and technology; and advance international cooperation in science.

Information for Authors
See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

EDITOR-IN-CHIEF **Bruce Alberts**
EXECUTIVE EDITOR **Monica M. Bradford**
DEPUTY EDITORS **R. Brooks Hanson, Barbara R. Jasny, Katrina L. Kelner**
NEWS EDITOR **Colin Norman**

Editorial supervisory senior editor Phillip D. Szuromi; **senior editor/perspectives** Lisa D. Chong; **senior editors** Gilbert J. Chin, Pamela J. Hines, Paula A. Kiberstis (Boston), Marc S. Lavine (Toronto), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Valda Vinson; **associate editors** Jake S. Yeston, Laura M. Zahn; **online editor** Stewart Wills; **associate online editors** Robert Frederick, Tara S. Marathe; **web content developer** Martyn Green; **book review editor** Sherman J. Suter; **associate letters editor** Jennifer Sills; **editorial manager** Cara Tate; **senior copy editors** Jeffrey E. Cook, Cynthia Howe, Harry Jach, Barbara P. Ordway, Trista Wagoner; **copy editors** Chris Filiatreau, Lauren Kmec, Peter Mooreside; **editorial coordinators** Carolyn Kyle, Beverly Shields; **publications assistants** Ramatoulaye Diop, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Scott Miller, Jerry Richardson, Jennifer A. Seibert, Brian White, Anita Wynn; **editorial assistants** Carlos L. Durham, Emily Guise, Patricia M. Moore; **executive assistant** Sylvia S. Kihara; **administrative support** Maryrose Madrid

News deputy news editors Robert Coontz, Eliot Marshall, Jeffrey Mervis, Leslie Roberts; **contributing editors** Elizabeth Culotta, Polly Shulman; **news writers** Yudhijit Bhattacharjee, Adrian Cho, Jennifer Couzin, David Grimm, Constance Holden, Jocelyn Kaiser, Richard A. Kerr, Eli Kintisch, Andrew Lawler (New England), Greg Miller, Elizabeth Pennisi, Robert F. Service (Pacific NW), Erik Stokstad; **interns** Rachel Zelkowitz, Andrea Lu, Fayana Richards; **contributing correspondents** Jon Cohen (San Diego, CA), Daniel Ferber, Ann Gibbons, Mitch Leslie, Charles C. Mann, Virginia Morell, Evelyn Strauss, Gary Taubes; **copy editors** Linda B. Felaco, Melvin Gatling; **administrative support** Scherraine Mack, Fannie Groom; **bureaus** New England: 207-549-7755, San Diego, CA: 760-942-3252, FAX 760-942-4979, Pacific Northwest: 503-963-1940

Production director James Landry; **senior manager** Wendy K. Shank; **assistant manager** Rebecca Doshi; **senior specialists** Steve Forrester, Chris Redwood; **specialist** Anthony Rosen; **Preflight director** David M. Tompkins; **manager** Marcus Spiegler; **specialist** Jessie Mudjitaba

Art director Kelly Buckheit Krause; **associate art director** Aaron Morales; **illustrators** Chris Bickel, Katharine Sutliff; **senior art associates** Holly Bishop, Laura Creveling, Preston Huey, Nayomi Kevitiyagala; **associate** Jessica Newfield; **photo editor** Leslie Blizard

Science International

Europe (science@science-int.co.uk) **editorial: international managing editor** Andrew M. Sugden; **senior editor/perspectives** Julia Fahrenkamp-Uppenbrink; **senior editors** Caroline Ash, Stella M. Hurtley, Ian S. Osborne, Peter Stern; **editorial support** Deborah Dennison, Rachel Roberts, Alice Whaley; **administrative support** John Cannell, Janet Clements, Louise Smith; **news: europe news editor** John Travis; **deputy news editor** Daniel Clery; **contributing correspondents** Michael Balter (Paris), John Bohannon (Vienna), Martin Enserink (Amsterdam and Paris), Gretchen Vogel (Berlin); **intern** Lauren Cahoon

Asia Japan Office: Asca Corporation, Eiko Ishioka, Fusako Tamura, 1-8-13, Hirano-cho, Chuo-ku, Osaka-shi, Osaka, 541-0046 Japan; +81 (0) 6 6202 6272, FAX +81 (0) 6 6202 6271; asca@os.gulf.or.jp; **asia news editor** Richard Stone (Beijing: rstone@aaas.org); **contributing correspondents** Dennis Normile (Japan: +81 (0) 3 3391 0630, FAX +81 (0) 3 5936 3531; dnormile@gol.com); Hao Xin (China: +86 (0) 10 6307 4439 or 6307 3676, FAX +86 (0) 10 6307 4358; cindyhao@gmail.com); Pallava Bagla (South Asia: +91 (0) 11 2271 2896; pbagla@vsnl.com)

Africa Robert Koenig (contributing correspondent, rob.koenig@gmail.com)

EXECUTIVE PUBLISHER **Alan I. Leshner**
PUBLISHER **Beth Rosner**

Fulfillment Systems and Operations (membership@aaas.org); **director** Waylon Butler; **senior systems analyst** Jonny Blaker; **customer service supervisor** Pat Butler; **specialists** Latoya Casteel, LaVonda Crawford, Vicki Linton; **data entry supervisor** Cynthia Johnson; **specialist** Tarrika Hill

Business Operations and Administration director Deborah Rivera-Wienhold; **assistant director, business operations** Randy Yi; **manager, business analysis** Michael LoBue; **manager, business operations** Jessica Tierney; **financial analysts** Benjamin Aronin, Priti Pamnani; **rights and permissions: administrator** Emilie David; **associate** Elizabeth Sandler; **marketing director** John Meyers; **marketing managers** Allison Pritchard, Darryl Walter; **marketing associates** Aimee Aponte, Alison Chandler, Mary Ellen Crowley, Marcia Leach, Julianne Wielga, Wendy Wise; **international marketing manager** Wendy Sturley; **marketing executive** Jennifer Reeves; **marketing/member services executive** Linda Rusk; **site license sales director** Tom Ryan; **sales manager** Russ Edra; **sales and customer service** Iquo Edim, Kiki Forsythe, Catherine Holland, Phillip Smith, Philip Tsolakidis; **electronic media: manager** Lizabeth Harman; **project manager** Trista Snyder; **assistant manager** Lisa Stanford; **senior production specialists** Christopher Coleman, Walter Jones; **production specialists** Nichele Johnston, Kimberly Oster

Advertising director worldwide ad sales Bill Moran

Product (science_advertising@aaas.org); **midwest** Rick Bongiovanni: 330-405-7080, FAX 330-405-7081; **west coast/w. canada** Teola Young: 650-964-2266; **east coast/e. canada** Laurie Faraday: 508-747-9395, FAX 617-507-8189; **uk/europe/asia** Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532; **japan** Mashy Yoshikawa: +81 (0) 3 3235 5961, FAX +81 (0) 3 3235 5852; **senior traffic associate** Deiandra Simms

Commercial Editor Sean Sanders: 202-326-6430

Classified (advertise@sciencecareers.org); **us: recruitment sales manager** Ian King: 202-326-6528, FAX 202-289-6742; **inside sales manager: midwest/canada** Daryl Anderson: 202-326-6543; **inside sales representative** Karen Foote: 202-326-6740; **key account manager** Joribah Able; **northeast** Alexis Fleming: 202-326-6578; **southeast** Tina Burks: 202-326-6577; **west** Nicholas Hintibidze: 202-326-6533; **sales coordinators** Erika Foard, Rohan Edmonson, Shirley Young; **international: sales manager** Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532; **sales** Marium Hudda, Alex Palmer, Alessandra Sorgente; **sales assistant** Louise Moore; **japan** Mashy Yoshikawa: +81 (0) 3 3235 5961, FAX +81 (0) 3 3235 5852; **advertising production operations manager** Deborah Tompkins; **senior production specialists** Robert Buck, Amy Hardcastle; **senior traffic associate** Christine Hall; **publications assistant** Mary Lagnaoui

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SENIOR EDITORIAL BOARD

BOARD OF REVIEWING EDITORS

BOOK REVIEW BOARD

End of a Tiger's Tale

The saga of faked photos of a wild South China tiger (*Science*, 14 December 2007, p. 1701) reached its denouement last month, when the photographer was arrested and 13 Shaanxi Province officials, including its top wildlife official, were sacked.

Last October, a farmer named Zhou Zhenglong produced 71 purported images of the reputedly extinct tiger at a much-touted press conference held by the province's forestry association (SFA). SFA gave him a 20,000 yuan ($2666) reward and began plans for a tiger reserve. But Chinese netizens soon pounced on the images, arguing that they were fakes. A national ruckus ensued and led to a full-scale criminal inquiry. Police found, among other things, that the "trees" in Zhou's photos were in reality only 0.8 centimeters wide.

At the 29 June press conference, Bai Shaokang, a spokesperson for Shaanxi's Public Security Bureau, announced that Zhou admitted he had cut a tiger picture from a calendar and stuck it in a bushy area to photograph it. He also used a carved wooden tiger paw to leave prints nearby. Zhou is being held on charges of swindling the government, and SFA has demanded that he return his reward.

The sorry episode has given the Chinese a new saying: "Zhenglong paihu"—meaning that something is as unbelievable as a tiger photo by Zhenglong.

The wave.

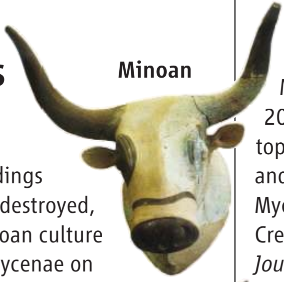
SOCCER SOLITON

A Rotterdam soccer stadium will briefly become a giant physics lab on 19 July, when Dutch researchers plan to carry out experiments to study the "wave," the ripple that races through a crowd as fans briefly stand up and raise their arms. Fluid dynamicist GertJan van Heijst of Eindhoven University of Technology in the Netherlands wants to test the idea that the wave is a soliton, a single wave that keeps its shape and travels at constant speed.

Solitons have a peculiar trait: When two of them collide, both emerge and continue on unchanged. If stadium waves are solitons, the same should happen there. To find out, Van Heijst wants to create colliding waves three times during festivities for the 100th anniversary of Feyenoord soccer club. A sports commentator will give the expected 50,000 fans slightly different instructions each time as cameras record what happens.

Scientists have never done field experiments with the wave, says Illés Farkas of the Hungarian Academy of Sciences in Budapest, who studies the phenomenon through analyses of videos and computer models. "This is really fascinating." The stunt is part of the bicentennial celebration of the Royal Netherlands Academy of Arts and Sciences.

Cretan Bones



Minoan

Mycenaean

About 3500 years ago, ancient Crete fell apart. Palaces and public buildings all over the island were destroyed, and the indigenous Minoan culture fell under the sway of Mycenae on the Greek mainland.

For years, experts pinned much of the seemingly abrupt changes on a Mycenaean invasion. But a recent analysis of bones in Cretan tombs indicates that the "invasion" was actually a local insurrection. Aegean scholars began to doubt the invasion theory some time ago, when excavations showed that the destruction had been selective and "Mycenaeanization" had been gradual, not sudden.

New support for a revised scenario comes from Argyro Nafplioti of the American School of Classical Studies in Athens. Nafplioti sampled dental enamel and thighbone from 30 individuals who had been buried near Knossos in Minoan graves and Mycenaean-style tombs before and after the 20-year period of destruction. Strontium isotope analysis and comparisons with ancient and modern animal tissues from Crete and Mycenae revealed that all were native-born Cretans, the researchers report in the August *Journal of Archaeological Science*. Archaeologist A. Bernard Knapp of the University of Glasgow, U.K., says the new analysis offers "a compelling corrective" to those who still see Crete as "the domain of 'Mycenaean' elites controlling [its] social, political, ideological, and material cultural traditions."

Horny Young Devils

A highly contagious facial cancer in Tasmanian devils (*Science*, 18 February 2005, p. 1035) has sparked a trend toward adolescent pregnancies in the endangered animals. The disease, which emerged about 10 years ago, strikes mostly adults and kills them within months. Juveniles seem to be stepping in to fill their reproductive role, Australian biologists say.

The scientists compared the average age at reproduction for devils at the same sites in Tasmania before and after the cancer's onset. Healthy devils typically start breeding at age 2 and have multiple litters. But devils at four out of the five sites tested began breeding by age 1 and typically produced only one litter before falling ill, the scientists reported online 14 July in the *Proceedings of the National Academy of Sciences*.

It's unclear whether this is a true case of evolution or if young devils are simply maturing earlier thanks to less competition for resources, says evolutionary biologist Nelson Hairston of Cornell University. Similar changes in reproductive patterns have been observed in fish and in mammals such as rabbits. This example adds "a very sexy example in a charismatic megafauna," he says.

CREDITS (TOP TO BOTTOM): FRANCK SEGUIN/TEMPSPORT/CORBIS; GLYN GENIN/DPA/CORBIS; THE BRIDGEMAN ART LIBRARY/GETTY IMAGES; COURTESY OF MENNA JONES

Published by AAAS

NEWSMAKERS

EDITED BY YUDHIJIT BHATTACHARJEE

Three Q's >>

German neuroscientist **Tobias Bonhoeffer** has been selected as the first president of the Institute of Science and Technology Austria. The graduate institute, scheduled to open next year in Klosterneuburg, northwest of Vienna, has been promised $860 million over 10 years in state funding. Bonhoeffer, who studies the cellular basis of learning and memory, is currently a director at the Max Planck Institute of Neurobiology in Martinsried, Germany.

Q: What got you interested in this job?
Over the last couple of years, I've been interested in making an impact on a bigger scale. It's an opportunity where one can really shape something. And Vienna is a very nice city.

Q: The institute's supporters initially said they were hoping for an "Austrian MIT." Do you see MIT as the best model for IST Austria?
We want an MIT, but only in terms of quality, not in terms of breadth. It is not possible to create an MIT from scratch. One will have to think hard and strategically about disciplines where we can be world-class and make a difference and those where it would be hard to compete. … I would like to give Austria something to be proud of—as proud as they are of the men's downhill ski team.

Q: Will you be able to do any science?
I would like to be able to organize things in such a way that I would not need to leave science completely. I also have to think about the future. I'm 48, so even if I were to stay for 12 years, I won't be retiring. And I don't want to have to start from scratch at 60.

Got a tip for this page? E-mail people@aaas.org

MONEY MATTERS

FOR THE PLANET. A 1952 Princeton alum has donated $100 million to the university for a research center to address energy and the environment.

Gerhard Andlinger, who grew up in Austria, is the chair of a global investment company that in recent years has invested in clean and renewable energy companies. The gift is intended to support research aimed at finding technological solutions to environmental problems. It will also fund the construction of a 10,219-square-meter engineering laboratory, basic geoscience research, and positions for faculty focused on environmental policy issues.

"My hope in establishing this center is to focus [Princeton's strengths] on finding 'clean-tech' solutions to the most important problems facing our society today," says Andlinger, according to a Princeton press release. "The work of the center will help create a better world for our children and grandchildren, which I see as a personal as well as institutional responsibility."

DEATHS

ON A QUEST. John Templeton, the U.S.-born philanthropist who supported research into what he called "spiritual realities," died 8 July at his home in the Bahamas. He was 95.

The foundation he established in 1987 provides about $60 million annually in grants for conferences and research on the origin and benefits of religion, the mechanisms behind concepts such as forgiveness and love, and other topics ranging from consciousness to cosmology. The $1.1 billion foundation also administers an annual $1.4 million prize that has honored the work of several scientists.

Some researchers have criticized Templeton's efforts to promote a convergence of science and religion; for example, University of Oxford biologist and avowed atheist Richard Dawkins calls the Templeton prize a "Faustian bargain." But many have welcomed Templeton's funding of these topics. "Whether or not we share [Templeton's] vision that religion and science will be reconciled, we must agree with him [that] they cannot continue to ignore each other," says William Bainbridge, a sociologist in the computing directorate at the U.S. National Science Foundation in Arlington, Virginia, who has reviewed grant proposals for the foundation.

IN BRIEF

The former science director of the Texas Education Agency (TEA) has filed a lawsuit accusing the agency of having violated the separation of church and state by adopting a "neutral" position on creationism. **Christina Comer**, who was fired from her job in November 2007 after forwarding an e-mail announcing a talk by a critic of the intelligent design movement (*Science*, 14 December 2007, p. 1703), is demanding that she be reinstated. The suit, filed 1 July in the U.S. District Court in Austin, argues that TEA's policy of neutrality is "not neutral at all, because it has the purpose or effect of inviting dispute about an issue—teaching creationism as science in public schools—that is forbidden by the Establishment Clause."

Movers

BACK TO ROOTS. Another star stem cell researcher is on the move—but this time not to California. John Gearhart, on the faculty at Johns Hopkins University in Baltimore, Maryland, since 1980, will head the University of Pennsylvania's new Institute for Regenerative Medicine, formed last November. Gearhart has done pioneering work with human pluripotent stem cells and is a high-profile advocate for loosening the Bush Administration's restrictions on federal funding for human embryonic stem cell research.

The move will be a homecoming for Gearhart, who was raised in an orphanage in Philadelphia after the death of his coal miner father. Gearhart's wife, geneticist Shannon Fisher, will also join the Penn faculty. Ralph Brinster and Jonathan Epstein have been interim co-directors of the institute.

CREDITS (TOP TO BOTTOM): IST; AP/MARTY LEDERHANDLER; MARTY KATZ ©2005

Published by AAAS

NEWS>>

THIS WEEK

CLIMATE CHANGE

Bush Takes a Final Swipe, and Salute, at CO_2 Emission Curbs

Floating target. Protesters at the G8 meeting in Japan decry ballooning U.S. and Canadian greenhouse gas emissions.

In the end, he just couldn't commit. Last week, the Bush Administration essentially ended its tumultuous relationship with climate change, unveiling two decisions that all but ensure that President George W. Bush will leave office without making a binding commitment to cut greenhouse gas emissions.

On 9 July, Bush and the other leaders of the Group of Eight (G8) industrial powers signed a largely symbolic pledge to help trim global emissions by 2050, rejecting stricter language. Then, on 11 July, the Administration announced that it would not use the nation's leading clean-air law to regulate heat-trapping gases, effectively sidestepping a U.S. Supreme Court decision.

Analysts say the two moves are probably the Administration's last gasp on climate. "These are the final major gestures; there's not much left for them to do. Now everybody's focused on what Congress and the next president will do," says Jody Freeman, head of the Environmental Law Program at Harvard Law School.

Both announcements reflect the fierce internal disagreements that have become hallmarks of the Administration's climate policy. As a presidential candidate in 2000, Bush backed using the Clean Air Act to regulate greenhouse gases. But he quickly backpedaled after winning office. State and local officials continued to press for action, however, arguing that carbon dioxide was a "pollutant" covered by the law (*Science*, 8 September 2006, p. 1375). And last year their arguments prevailed, when the Supreme Court ordered the Environmental Protection Agency (EPA) to explain why it wasn't regulating the gas.

The Administration split over how to respond. One faction, led by senior EPA officials, drafted a detailed rationale for using the law to attack climate change. But that road map drew furious objections from Vice President Dick Cheney and other White House officials, including presidential science adviser John Marburger, according to documents released by EPA. A quartet of Cabinet members also chimed in, according to EPA; the secretaries of Agriculture, Transportation, Commerce, and Energy complained that it did not "fairly recognize the enormous ▶

RESEARCH ETHICS

Old Samples Trip Up Tokyo Team

TOKYO—A University of Tokyo team has retracted a published research paper because it apparently failed to obtain informed consent from tissue donors or approval from an institutional review board (IRB). Other papers by the same group are under investigation by the university. Observers believe problems stem in part from guidelines that don't sufficiently explain how to handle samples collected before Japan established informed consent procedures.

The alleged infractions were announced at a press conference on 11 July by Motoharu Seiki, dean of the university's Institute of Medical Science (IMS). Seiki did not identify the researchers, but *Asahi Shimbun*, a prominent daily, broke the story the morning of the press conference and reported that the authors are members of a group led by Arinobu Tojo, who works on molecular therapies for leukemia. No one answered Tojo's office phone, and he did not immediately respond to an e-mail from *Science*.

The withdrawn paper was published online on 21 May and in the 1 July issue of *Haematologica*. A statement on the journal's Web site says a paper on acute myeloid leukemia by Seiichiro Kobayashi *et al*. was retracted on 27 June by Tojo, the corresponding author, who had informed the editors that an investigation found that the "study had not been approved by the IRB."

Seiki says tissue samples used for the retracted paper were collected before Japan's Ministry of Health issued guidelines for IRBs and informed consent in 2003. IMS's policies call for an IRB review of the use of old samples. "But [the researchers] did not follow that process," says Seiki. Tohru Masui, a bioresources policy specialist at the National Institute of Biomedical Innovation in Osaka, says few researchers are aware of the ethical issues surrounding old samples because the ministry guidelines "do not have [anything] about legacy samples."

Seiki says an external review panel has been established and will report its findings by the end of this month. **–DENNIS NORMILE**

CREDIT: REUTERS/KIM KYUNG-HOON/LANDOV

Building
better rice
330

Model sleep
organisms
334

and, we believe, insurmountable burdens, difficulties, and costs" of the strategy.

EPA chief Stephen Johnson told reporters that the infighting convinced him that "the Clean Air Act is the wrong tool for the job" and that it would be impossible to forge a consensus response "in a timely manner." Instead, he issued a 588-page document that laid out dozens of complex questions raised by the law. The document also laid bare the squabbling and asked the public to join the debate. Johnson said he hopes the move will convince Congress that entirely new laws are needed to deal with climate change.

That's probably true, says Kevin Vranes, a former Senate staffer now working with Point 380—the name refers to the current level of carbon dioxide in the atmosphere, in parts per million—in Boulder, Colorado. Given the lack of White House leadership, he says that "Congress needs to stop stalling … and start addressing the problem itself." Both major presidential candidates, senators Barack Obama (D–IL) and John McCain (R–AZ), have embraced some sort of controls on greenhouse gases, although in May the Senate handily rejected a plan to do so by means of a cap-and-trade system.

The G8 declaration to seek a 50% reduction in emissions includes no interim targets and no mechanism for achieving the goal. But it may polish Bush's legacy by pointing toward a new global climate deal in 2009. The Administration could still do "potentially very constructive work" on the global stage, says David Victor, director of an energy and development program at Stanford University in Palo Alto, California. But those efforts may not "have much lasting power, since nearly all the rest of the world is already looking beyond Bush."

–DAVID MALAKOFF

David Malakoff is a science writer in Alexandria, Virginia.

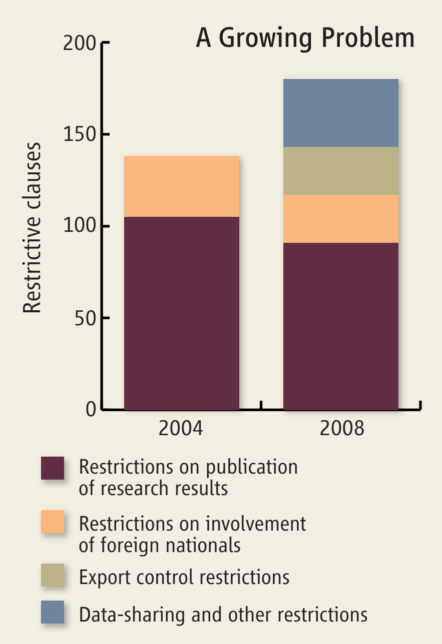
DEFENSE RESEARCH

New Policy Tries to Ease Security Restrictions

The grant that the U.S. Army Corps of Engineers awarded Zdenek Bazant of Northwestern University earlier this year to study how tough materials are able to withstand impact came with a catch: The corps had to vet any foreign nationals working on the project. Officials at the Evanston, Illinois, university balked, saying the requirement violates the school's antidiscrimination policies. Now the university has a new argument: The restriction also contradicts a new policy directive from the corps's parent agency, the Department of Defense (DOD), that's meant to resolve a 7-year dispute between the Pentagon and academic institutions over the rules governing unclassified research.

Since the terrorist strikes of 11 September 2001, research agencies have tried to prevent sensitive technical information from falling into enemy hands by creating a category known as "sensitive but unclassified" research. Academic officials have fought back, pointing to a 1985 directive from the Reagan Administration that exempts fundamental research on university campuses from such restrictions. Last month, the universities won a major victory when DOD Under Secretary John Young instructed agency officials that "classification is the only appropriate mechanism" for restricting publications or participation of foreign nationals in unclassified research projects. "The performance of fundamental research, with rare exceptions, should not be managed in a way that it becomes subject to restrictions on the involvement of foreign researchers or, publication restrictions," the memo says, citing National Security Defense Directive 189, which President Ronald Reagan issued.

"We felt that there was a need to remind everyone" that fundamental research is to remain free of restrictions, says William Rees, the Pentagon's head of basic research, who led the internal review. "The strength of American science demands a research environment that is fully conducive to the free exchange of ideas."



Closely guarded. The number of restrictions on research contracts has grown since the first survey of 20 universities in 2004.

SOURCE: AAU-COGR

That hasn't been the case, says a report last year by the National Academies' National Research Council. A survey of more than 20 universities by the Association of American Universities (AAU) and the Council on Government Relations documents 180 instances of troublesome clauses in research contracts from federal agencies, a majority from DOD and the Department of Homeland Security (see graphic). So the new policy is a welcome change, says Jacques Gansler, a former Pentagon administrator who co-chaired the academies' report.

"We are very pleased with the directive," says Gansler, now a professor at the University of Maryland, College Park, who hopes that other federal agencies will follow DOD's example. AAU's Toby Smith says he had hoped Young would also ban companies from passing along such restrictive language to university subcontractors, but he's glad the memo asks DOD authorities to retrain the agency's contracting officers.

Such training seems essential, say Northwestern officials, who are still negotiating with the Army over Bazant's award after a corps official said Young's memo did not invalidate the corps' own policies. "We may have to decline the award," a university official told *Science*.

–YUDHIJIT BHATTACHARJEE

Spokeswoman. Actress Jenny McCarthy (*center*) has described on talk shows and at rallies, such as this one held in Washington, D.C., in June, how chelation helped her son recover from autism.

CREDIT: MANDEL NGAN/AFP/GETTY IMAGES

MEDICINE

Stalled Trial for Autism Highlights Dilemma of Alternative Treatments

The tension between parents desperate to help their sick children and researchers who worry about quack medicine has long put public health agencies in a bind. Last week, a long-simmering controversy boiled over when newspapers across the country ran an Associated Press story claiming that "government researchers are pushing to test an unproven treatment on autistic children, a move some scientists see as an unethical experiment in voodoo medicine." In fact, a trial of the controversial treatment was halted last year, and Thomas Insel, director of the National Institute of Mental Health (NIMH) in Bethesda, Maryland, says he's not pushing to restart it. The case, and the publicity surrounding it, illustrates the difficulty of deciding whether to test these questionable therapies, especially in children.

The "voodoo" here is chelation therapy. Believing that mercury in vaccines triggers autism, thousands of parents, often at the advice of their physicians, have given their autistic children drugs to bind, or chelate, and remove heavy metals from the body. Some say the over-the-counter or off-label treatment can improve poor language skills, social problems, and other symptoms of the disorder. And yet the drugs are not risk-free, and the underlying rationale—that mercury from vaccines causes or worsens autism—has been roundly rejected by many scientific studies.

NIMH has argued that the widespread use of the drugs creates a "public health imperative" to conduct a rigorous trial so that the institute can inform parents and physicians about any merits or dangers of the drugs. But some researchers and ethicists oppose studies that they say have no chance of working—and little chance of persuading the most zealous advocates—especially if the drug poses a substantial risk. "On balance, it's not an ethical study," says vaccine researcher Paul Offit of the University of Pennsylvania.

This isn't the first time researchers at the National Institutes of Health (NIH) have felt compelled to react to the use of dubious autism treatments. In the late 1990s, a wave of media publicity touted the abilities of a gastrointestinal drug called secretin to "cure" symptoms of autism (*Science*, 5 October 2001, p. 37). So many parents were buying the drug that NIH decided to do a series of small, rapid clinical trials. Secretin flopped, and most parents eventually stopped clamoring for it. "It was a meteoric rise, and it fell just as quickly," says one autism researcher, who asked not to be named to avoid offending advocates. "I haven't heard of anyone using secretin in years."

Chelation therapy remains widely used. Some surveys have suggested that 2% to 8% of children with autism have had it, perhaps several thousand per year. Parents either buy unregulated supplements or have a doctor use a treatment for lead poisoning. Not only do the drugs bind to toxic metals, but they can also remove essential minerals such as calcium and iron.

NIMH wanted to conduct a study of the common chelator DMSA, which is approved by the Food and Drug Administration for treating lead poisoning. The idea was to give 120 children, aged 4 to 10, with a range of autism symptoms either DMSA or a placebo. After 12 weeks, NIMH researchers would evaluate the children to see if their social and language skills had improved. It would be the first controlled study of a chelator on autism.

But first the study had to pass ethical muster with a so-called institutional review board (IRB). Putting children at risk of side effects is considered unethical if they are unlikely to receive any direct benefit from the drug. And Insel acknowledges that "it is difficult to make the case" that a chelator would help children with autism. On the other hand, a well-conducted trial with negative results could help parents better choose whether to use a chelator, says pediatrician and bioethicist Douglas Diekema of the University of Washington, Seattle, who was not part of the IRB. The NIMH study, which included multivitamins to safeguard against most of the risks of the drug, passed review and was launched in September 2006.

A few months later, new research raised a red flag. An October 2006 online study in *Environmental Health Perspectives* examined the impact of DMSA on rodents. Although the drug helped rodents overcome lead poisoning, when it was given to rodents without lead it caused lasting cognitive and emotional problems. The finding "raises concerns about the use of chelating agents in treating autistic children without elevated levels of heavy metals," says senior author Barbara Strupp of Cornell University, although she notes that it's not known what the threshold might be for such adverse effects. The children in the autism trial would not have had elevated levels of mercury in their blood (otherwise, they could not ethically be given a placebo).

NIMH officials halted the trial in February 2007 and sent it back to the IRB for further review. Given the new risks, the IRB concluded it did not have the authority to approve the trial, although NIMH's parent agency, the Department of Health and Human Services (HHS), could if it felt the societal benefit were large enough. Rather than appeal to HHS, Insel says, the principal investigator, NIMH's Susan Swedo, decided that NIMH's intramural resources were better focused elsewhere, on the possible benefit of reducing inflammation with an antibiotic called minocycline in children with so-called regressive autism. Some critics of the chelation therapy say it was a good call because there is some preliminary evidence to suggest why inflammation—as opposed to mercury—might be involved in autism.

–ERIK STOKSTAD

PALEOCEANOGRAPHY

Caribbean Megaeruptions Drove a Global Ocean Crisis

Things got pretty ugly in the world ocean 93.5 million years ago. Deeper waters turned foul as their oxygen disappeared and the sea floor around the globe became a lethal black ooze. Many bottom-dwelling shelled animals from the microscopic to the gigantic went extinct. Now new geochemical evidence recovered from that ancient muck strongly links this global crisis—called Oceanic Anoxic Event 2 (OAE2)—to one of the world's largest episodes of volcanism.

The new work "nails the coffin shut" on this long-suspected volcanic connection, says paleoceanographer Timothy Bralower of Pennsylvania State University in State College. The finding also adds support to nearly a half-dozen other proposed volcanic crises during the past 250 million years, including the greatest mass extinction of them all.

OAE2 "was the big one," says Bralower, who was not involved in the new work. "It was the most global, the most dramatic" of a half-dozen OAEs during the exceptional warmth of the mid-Cretaceous period 120 million to 80 million years ago. The young Atlantic Ocean was as narrow as a few hundred kilometers, the sea ran free between Europe and Africa and into the western Pacific, and high sea levels drove the ocean up onto the continents.

Something in this mid-Cretaceous world had made the ocean liable to shift dramatically the way it operated. During OAE2 about 93.5 million years ago, for example, life-giving oxygen abruptly disappeared from deeper waters, and so much organic matter accumulated in muddy bottom sediments that for a half-million years the sediment turned black until the seas recovered. Paleoceanographers looking for triggers of OAEs, especially OAE2, have long turned their attention to humongous volcanic eruptions, such as the lava outpourings of a large igneous province (LIP) now lying beneath the Caribbean Sea. A shift in lead isotopes recorded at the very onset of OAE2 in Italy supported that idea (*Science*, 27 April 2007, p. 527), but the evidence remained regional in scale.

This week in *Nature*, paleoceanographers Steven Turgeon and Robert Creaser of the University of Alberta (UA) in Edmonton, Canada, report geographically broad-based isotopic evidence for a volcano-OAE2 link. They measured the element osmium in sediments across OAE2 from Italy—which was in the Tethys seaway between Europe and Africa at the time—and just off northeast South America, which was then in the opening Atlantic.

At both sites, the osmium abundance shot up by a factor of 30 to 50 above background just before the onset of OAE2. In the Atlantic, the lag between osmium increase and anoxia was between 10,000 and 20,000 years, the UA researchers estimate. And just as vastly more osmium was entering the ocean, the ratio of osmium-187 to osmium-188 plummeted. All that is just what would happen, say Turgeon and Creaser, when thousands upon thousands of cubic kilometers of lava delivered osmium from Earth's mantle to the sea floor of the Caribbean, a LIP eruption previously dated to within a few million years of OAE2.

Crunch time. A dark band in an Italian quarry marks an ocean crisis 93.5 million years ago.

CREDIT: STEVEN TURGEON

The new osmium data "do make the argument more compelling" that the largest eruptions can trigger anoxic crises in the ocean, says Millard Coffin of the University of Southampton, U.K., who specializes in LIPs. The trigger "is most likely volcanic," he agrees. The work has broader implications too. The largest LIP of the past half-billion years—the Siberian Traps—seems to have coincided with the largest mass extinction, the Permian-Triassic, but dating uncertainties still allow the extinctions to precede the eruptions by hundreds of thousands of years (*Science*, 25 April, p. 434). In the case of OAE2, at least, the coincidence was tighter still. **–RICHARD A. KERR**

SCIENCE**SCOPE**

Occupational Safety Proves an Unsafe Occupation

U.S. health researchers are worrying about the future of federal research on worker safety following the puzzling decision to let the popular director of the National Institute for Occupational Safety and Health (NIOSH) go. Centers for Disease Control and Prevention (CDC) Director Julie Gerberding announced 3 July that she plans to replace NIOSH Director John Howard after his 6-year assignment ends this week despite his interest in serving another term. The physician had resounding support from labor, business, and health professionals. "It's really distressing. He's been a great guy," says Sarah Felknor of the University of Texas School of Public Health in Houston, who chairs NIOSH's board of scientific counselors. She and others are worried about the continuity of programs such as nanotoxicology research. Others fear that CDC will now push ahead with a plan that Howard resisted 4 years ago to move NIOSH down in the CDC hierarchy (*Science*, 16 July 2004, p. 323). A CDC spokesperson says that "there are no plans to reorganize NIOSH," and Gerberding was unavailable for comment. **–JOCELYN KAISER**

Tough New Conflicts Rules

Three leading journals have adopted or announced plans to adopt conflict-of-interest disclosure policies of unprecedented strictness. *Addiction*, the *Journal of the National Cancer Institute*, and the *Journal of the American College of Surgeons* are now requiring authors to disclose every financial tie, regardless of size, held within 3 years prior to submission. Editors from each journal collaborated with the Center for Science in the Public Interest (CSPI) to develop the policy, which the center released this week. *Addiction* editor Thomas Babor says it's an attempt to preserve scientific integrity amid revelations of scientists who concealed industry backing of their research. CSPI has circulated the model among hundreds of journals but no others have signed on; many feel their current policies are sufficient, says CSPI project director Merrill Goozner.

Others may fear the added burden such a stringent policy could impose on submitters. But Babor says the policy creates little extra work for the journals and only slightly more for authors—work that researchers are beginning to expect as par for the course of submitting a paper. "We are being perceived as a better journal," he says, noting steadily rising submissions. **–RACHEL ZELKOWITZ**

Published by AAAS

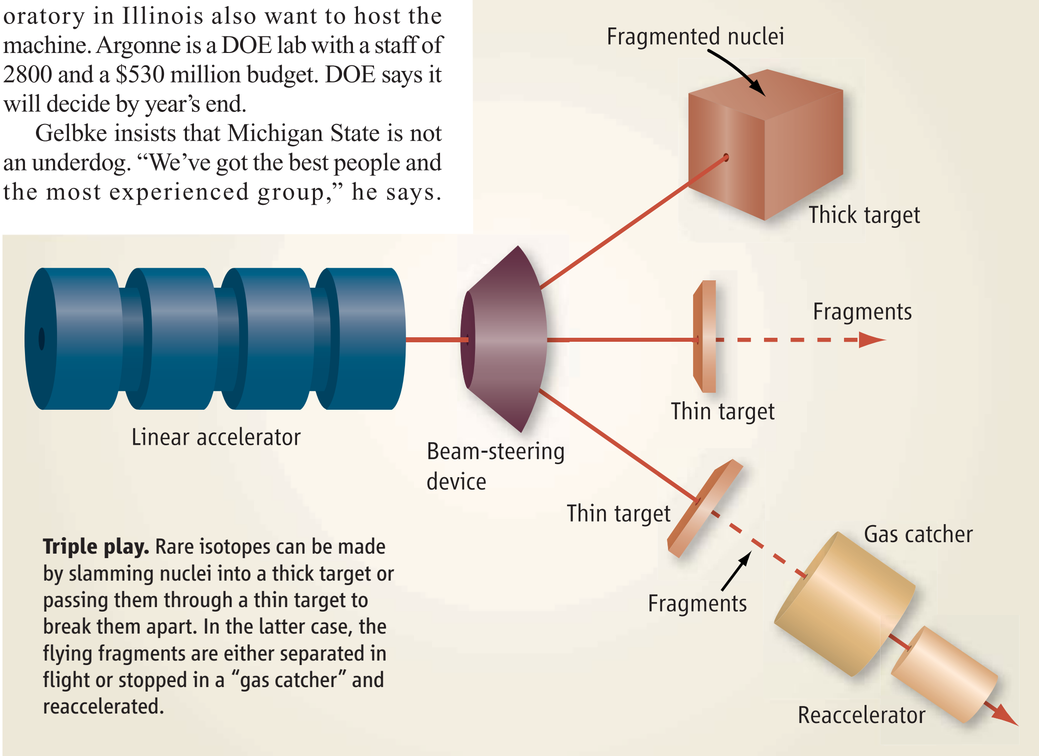
NUCLEAR PHYSICS

Two U.S. Labs Vie for Long-Delayed Exotic Nuclei Source

Can a small group of university researchers triumph over a big national laboratory in a competition to build and operate a $550 million piece of scientific machinery? C. Konrad Gelbke, a nuclear physicist at Michigan State University in East Lansing, and his colleagues are about to find out.

Next week, the U.S. Department of Energy (DOE) will accept proposals for a Facility for Rare Isotope Beams (FRIB), an accelerator to make fleeting nuclei never before produced outside stellar explosions. Gelbke and colleagues want to build FRIB at Michigan State's National Superconducting Cyclotron Laboratory (NSCL), a facility already pursuing such work with 300 employees and an annual budget of $20 million from the U.S. National Science Foundation (NSF). But researchers from Argonne National Laboratory in Illinois also want to host the machine. Argonne is a DOE lab with a staff of 2800 and a $530 million budget. DOE says it will decide by year's end.

Gelbke insists that Michigan State is not an underdog. "We've got the best people and the most experienced group," he says. "That's just established fact." But others who have observed similar competitions say Argonne's greater resources and existing infrastructure could give it a significant edge. "It is fundamentally an asymmetric situation, and it gets down to how important is that existing infrastructure?" says Michael Witherell, a particle physicist at the University of California, Santa Barbara, and former director of Fermi National Accelerator Laboratory in Batavia, Illinois.



Triple play. Rare isotopes can be made by slamming nuclei into a thick target or passing them through a thin target to break them apart. In the latter case, the flying fragments are either separated in flight or stopped in a "gas catcher" and reaccelerated.

FRIB is the second design of a machine that could reveal the birthplace of many heavy elements and hammer out a unified theory of nuclei large and small. Scientists know that more than half the elements heavier than iron originate somewhere in exploding stars through the so-called r-process, in which a light nucleus quickly absorbs many neutrons. FRIB would make some of the intermediary nuclei and help pin down exactly when and where within a stellar explosion the r-process takes place. More generally, it would help scientists weave a hodgepodge of theoretical models into a comprehensive understanding of the nucleus.

Researchers started planning for such a machine, originally dubbed the Rare Isotope Accelerator (RIA), in 1999. The heart of the machine is a superconducting linear accelerator that can accelerate any nucleus from hydrogen to uranium. With a price tag of $1 billion, RIA aimed to set new standards for every method of producing isotopes (see figure).

In 2003, RIA tied for third place in a list of 28 facilities DOE hoped to build in the following 20 years. The next year, it passed the first of five "critical decision" reviews. But in February 2006, Secretary of Energy Samuel Bodman unexpectedly announced that the project would be delayed at least 5 years (*Science*, 24 February 2006, p. 1082). A month later, DOE asked for a cheaper, more focused design.

The community responded with FRIB, whose linear accelerator will reach energies half as high as those planned for RIA. The new design also accentuates one means of making isotope beams, called reacceleration. Still, the machine would produce exotic nuclei at rates up to 100 times greater than those achieved by competing machines in Japan and Germany. The Michigan State and Argonne teams won't discuss their designs until they've been submitted, but in the past the teams have emphasized different secondary techniques in addition to reacceleration.

The current contest is the latest in which size may matter. In 1993, DOE chose its own Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, over Cornell University for the site of a particle smasher that would crank out particles called B mesons and study the asymmetry between matter and antimatter. Like Michigan State, Cornell had a smaller NSF-funded facility, and some accused DOE of nepotism. But even Karl Berkelman of Cornell says SLAC's resources contributed to the project's success. "I'm not sure we could have done as good a job," he says.

Michigan State's Bradley Sherrill says NSCL's track record proves it's up to the task. "We have already demonstrated that we can design, build, and operate a rare-isotope user facility," he says. For his part, Walter Henning, a nuclear physicist who leads Argonne's effort, says that "all factors should be considered."

In 2006, the NSF renewed Michigan State's grant for 5 more years, so the lab's immediate future is secure. But if it does not land FRIB, the lab may close once the new machine is turned on.

DOE hopes to begin construction in 2013 and finish within 5 years. But that depends on its budget. DOE has requested $7 million for design work in 2009, but it will be up to Congress and the next Administration to follow through on the project.

–ADRIAN CHO

ILLUSTRATION: L. CREVELING/SCIENCE

Published by AAAS

SOCIOLOGY

Survey Finds Citations Growing Narrower as Journals Move Online

Millions of scholarly articles have migrated online in recent years, making trips to library stacks mostly obsolete. How has this affected research, which depends on published work to guide and bolster academic inquiry? A sociologist at the University of Chicago in Illinois argues on page 395 that the shift has narrowed citations to more recent and less diverse articles than before—the opposite of what most people expected.

Working solo, James Evans of the University of Chicago was curious about how citation behavior has changed in the sciences and social sciences. In theory, online access should make it quicker and easier for researchers to find what they're looking for, particularly now that more than 1 million articles are available for free.

Relying on Thomson Scientific's citation indexes and Fulltext Sources Online, Evans surveyed 34 million articles with citations from 1945 to 2005. For every additional year of back issues that a particular journal posted online, Evans found on average 14% fewer distinct citations to that journal, suggesting a convergence on a smaller pool of articles. In other words, as more issues of a journal were posted online, fewer distinct articles from that journal were cited, although there were not necessarily fewer total references to that journal. It suggests herd behavior among authors: A smaller number of articles than in the past are winning the popularity contest, pulling ahead of the pack in citations, even though more articles than ever before are available. The average age of citations also dropped. Valuable papers might "end up getting lost in the archives," says Evans.

Oddly, "our studies show the opposite," says Carol Tenopir, an information scientist at the University of Tennessee, Knoxville. She and her statistician colleague Donald King of the University of North Carolina, Chapel Hill, have surveyed thousands of scientists over the years for their scholarly reading habits. They found that scientists are reading older articles and reading more broadly—at least one article a year from 23 different journals, compared with 13 journals in the late 1970s. In legal research, too, "people are going further back," says Dana Neacşu, head of public services at Columbia University's Law School Library in New York City, who has studied the question.

One possible explanation for the disparate results in older citations is that Evans's findings reflect shorter publishing times. "Say I wrote a paper in 2007" that didn't come out for a year, says Luis Amaral, a physicist working on complex systems at Northwestern University in Evanston, Illinois, whose findings clash with Evans's. "This paper with a date of 2008 is citing papers from 2005, 2006." But if the journal publishes the paper the same year it was submitted, 2007, its citations will appear more recent. Evans disputes that this affected his results, noting that in many fields, such as economics, the time to publication remains sluggish.

Tight focus. Citations to journals that have been online longer, according to James Evans, tend to cluster around more recent dates.

CREDIT: DAVE G. HOUSER/CORBIS

In other ways, Evans's findings reflect the efficiency that comes with online searching. "There's always been a desire to be focused in your citations, but it was impossible to manifest that in the old world," says Michael Eisen, a computational biologist at the University of California, Berkeley, who helped found the Public Library of Science.

In the end, Evans notes that "I don't have snapshots of people in their offices searching." But, he says, his findings show that "everyone's shifting to this central set of publications"—an effect that may lead to easier consensus and less active debate in academia.

–JENNIFER COUZIN

SCIENCESCOPE

New Money for New Neuroscience

A childhood friendship has blossomed into plans for the first privately financed Max Planck Institute. Twin brothers Andreas and Thomas Strüngmann, 58, announced this week that they have donated €200 million for a new cognitive neuroscience institute in Frankfurt, Germany, to be administered by the Max Planck Society. The twins, who made their fortune in pharmaceuticals, have had their interest in the brain fed by childhood pal Wolf Singer, who became a neuroscientist and then director of the Max Planck Institute for Brain Research in Frankfurt. Singer will serve as acting head of the Ernst Strüngmann Institute, named for the donors' father. The society will have full control over scientific aspects of the new institute, Singer says. **–GRETCHEN VOGEL**

Ready. Set. Fuse!

The world's fusion researchers now have a new toy to keep them busy over the next 10 years while the International Thermonuclear Experimental Reactor (ITER) is being built in Cadarache, France. The $420 million Korea Superconducting Tokamak Advanced Research (KSTAR) reactor in Daejeon, South Korea, achieved its first plasma last month, and this week officials formally announced it ready for use, with full operations to begin next year. Construction director Gyung-Su Lee says the new reactor is an ideal training ground for ITER because its superconducting magnets are made of the same niobium-tin alloy that ITER will use, so researchers can test fabrication techniques. **–DENNIS NORMILE**

No SLAC From Stanford

The Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, will change its name. That's because the Department of Energy (DOE) wants to trademark the names of its 17 national labs, and Stanford University won't let DOE claim its name. The impasse is "an example of DOE idiocy," says Nobelist and former SLAC Director Burton Richter, because the university would protect the name anyway. But DOE has been pushing to trademark lab names for more than a decade, spurred in part by a new law allowing trademark suits against the government, a risk DOE says it is trying to minimize. **–ADRIAN CHO**

Published by AAAS

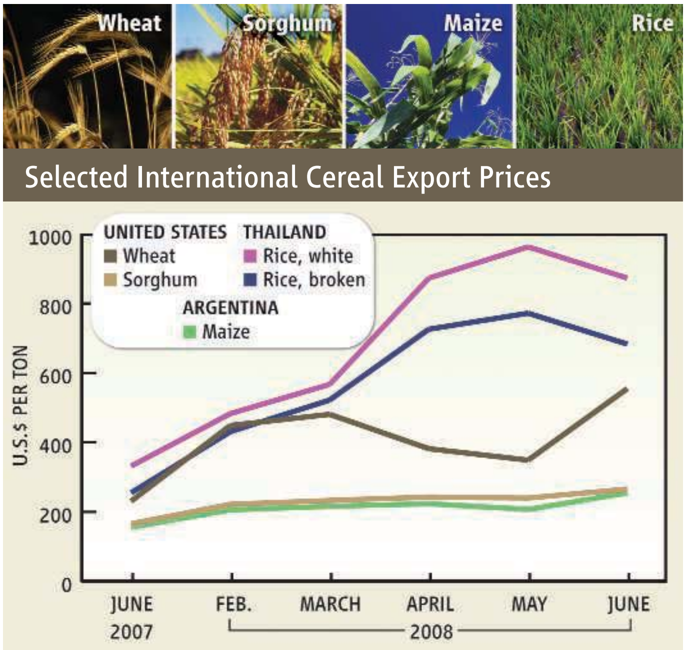
Reinventing Rice to Feed the World

With prices of rice and other cereals soaring and granaries emptying, it might take a second green revolution to avert widespread famine

SOUTH ASIA'S MONSOON IS A MIXED blessing for rice farmers. The rains fill paddies. Light flooding brings sediment that replenishes soil nutrients. But almost every year, somewhere, flooding is so severe it wipes out the crop.

In 2007, disaster struck the floodplains of the Tista and Jamuna rivers in north-central Bangladesh. Over a million hectares of farm fields were flooded, some inundated for as long as 3 weeks. Agricultural losses topped $600 million. A few pioneering farmers, however, were testing an experimental rice variety that tolerates submergence, and their plants recovered even after 12 days underwater—three times longer than normal varieties can endure. Yields suffered: They got about 4 tons per hectare, about 1 ton less than they would have without flooding, according to M. A. Salam, research director at the Bangladesh Rice Research Institute (BRRI) in Gazipur. "The farmers were very happy to get this yield under these circumstances," he says, because many of their neighbors were left with nothing.

Submergence-tolerant rice and other new yield-boosting varieties are arriving at a critical time. In recent weeks, the collision of rising demand and tightening supplies has driven a phenomenal spike in rice prices that sparked riots in Haiti, Bangladesh, and Egypt. A dozen countries, including India and China, have restricted rice exports, deepening the crisis. Exacerbating a bad situation, rice production in Myanmar this year will likely drop 6%, to 9.4 million tons, according to a U.S. Department of Agriculture (USDA) forecast, after extensive damage from Cyclone Nargis in early May. A storm surge flooded about 1.75 million hectares of the Irrawaddy River delta with saltwater and destroyed embankments and irrigation systems.



Selected International Cereal Export Prices

CREDITS: JUPITER IMAGES; SOURCE: FAO, GIEWS INTERNATIONAL CEREAL EXPORT PRICES, 20 JUNE 2008

The global food crisis grabbed the attention of G8 leaders meeting in Japan last week. They pledged to reverse the decline of aid and investment in agriculture and accelerate research and development (R&D) to boost food production. Nevertheless, the looming food shortage "is a story that's going to be here for a while," says Philip Pardey, an agricultural economist at the University of Minnesota, St. Paul. Demand will continue to rise, he says, as the world's population grows and more grain is diverted to produce biofuels and to feed livestock as meat consumption rises. At the same time, Pardey says, funding constraints have slowed R&D on improving grain yields and have crippled developing country extension systems, which get the latest seeds and techniques into farmers' hands.

All grains are affected by the trend. But a rice shortfall could be disastrous. In 2005, rice supplied 20% of total calories consumed worldwide, including 30% in Asia, according to the International Rice Research Institute (IRRI) in Los Baños, Philippines. IRRI claims that two-thirds of the world's poor—those living on less than $1 per day—subsist primarily on rice. And production is stagnant. Over the past several years, more rice has been consumed than grown—the difference made up by dipping into world rice stockpiles, which peaked at 146.7 million tons in 2001 but declined to 73.2 million tons in 2006, according to USDA. Prices were already rising, then lackluster harvests, export restrictions, and speculative buying sent prices soaring. For example, a popular export variety of Thai rice jumped from $362 per ton last December to $1000 per ton in April. Prices have retreated to $720 per ton.

To balance production and consumption, IRRI forecasts that by 2015 the world must grow 50 million tons more rice per year than the 631.5 million tons grown in 2005. This will require boosting global average yields by more than 1.2% per year, or about 12% over the decade, says IRRI's research director, Achim Dobermann. In the near term, he says, farmers could wring an extra 1 to 2 tons of grain per hectare by growing the latest varieties and improving farm management—everything from optimizing fertilizer use to building rat-proof granaries that stem

High and dry. An IRRI researcher examines submergence-tolerant rice that survived being underwater for 2 weeks; a nontolerant variety, to the left, perished.

CREDIT: ARIEL JAVELLANA/CPS/IRRI

postharvest losses. A long-term trial plot at IRRI produces 18 to 20 tons of grain per year per hectare, but the average field in Asia yields half of that. Existing technologies "haven't been moved out sufficiently to farmers," Dobermann says, because many extension systems are poorly funded and staffed. IRRI runs a training program that helps address this issue (see sidebar, p. 332).

In the long term, superior rice varieties are key to averting widespread food scarcity, says Pardey: "The yield levels we're seeing are historically high, and to even maintain them let alone increase them, you have to run pretty hard to keep ahead of evolving pests and diseases and other stresses." Given how long it takes to develop new varieties, he adds, "you've just got to keep priming the pump of the research."

No quick fix

Submergence-tolerant rice shows the years of effort it often takes to produce a new variety. Flooding costs South Asia about $1 billion a year in rice losses. Although paddy rice is grown in standing water, most varieties die if submerged for 3 or 4 days. Researchers had long known of varieties that apparently evolved to withstand monsoon flooding. An Indian variety known as FR (flood resistant) 13A can recover and produce rice even after 3 weeks underwater.

Despite that advantage, farmers have largely abandoned such varieties in favor of modern cultivars that produce double or more grain under normal conditions. In the 1970s, IRRI tried crossing FR13A with high-yield varieties. But farmers rejected the resulting cultivars because they didn't like the taste and had difficulty adapting the plants to local conditions, says David Mackill, head of plant breeding at IRRI.

In the early 1990s, Mackill, then at USDA's Agricultural Research Service in Davis, California, and colleagues at the University of California (UC), Davis, set out to identify the gene or genes in FR13A responsible for submergence tolerance. His team hybridized a variety derived from FR13A and an intolerant rice cultivar and tested hundreds of plants to see which recovered from submergence. Using molecular markers, or segments of easily identifiable DNA, they compared the genomes of the tolerant and nontolerant offspring, linking a region of chromosome 9 to submergence tolerance.

They enlisted colleagues at UC Riverside and IRRI to isolate the gene responsible, Submergence 1A (Sub1A). The group determined that Sub1A is expressed in FR13A only when the plant is submerged and that many nontolerant rice varieties don't have Sub1A. To confirm its role, they introduced Sub1A into an intolerant variety lacking the gene and got submergence tolerance. The group reported its findings in *Nature* in 2006.

IRRI plant physiologists, meanwhile, concluded that Sub1A inhibits stem and leaf elongation and the loss of chlorophyll that typically occurs in submerged plants. Limiting elongation conserves energy, and preserving chlorophyll, essential for photosynthesis, enhances chances of recovery.

Mackill joined IRRI in 2001 and 2 years later started working to get Sub1A into commercial varieties. Using marker-assisted selection, which links a DNA segment to a trait of interest, his team screened crosses for plants with Sub1A but otherwise identical to the target variety. The Swarna variety popular in India and Bangladesh was one of the first to get Sub1A, and germ plasm was given to BRRI and its counterpart in India in 2005. This year, BRRI has four varieties with the Sub1A gene in field trials, Salam says. They will ramp up seed production of the best candidate, which will take another 2 years. Varieties are being tested in eight other Asian countries. Production of submergence-tolerant rice will become appreciable sometime after 2010, Dobermann says.

It's fortunate that a single gene confers a high degree of submergence tolerance. Researchers aren't always so lucky. In 2002, a team at IRRI, the Philippine Rice Research Institute in Muñoz, BRRI, and UC Davis identified Saltol, short for salt tolerance, on rice chromosome 1. A rice variety carrying Saltol is now in field trials in Bangladesh. But Saltol confers tolerance only during the seedling stage. This works for wet-season rice, because adult plants are saved by monsoon rains that reduce soil salinity as the season progresses. But dry-season varieties face increasing salinity during the critical flowering period in spring, when coastal groundwater turns brackish. Researchers are probing for other genes that might protect these types.

Scientists are using molecular techniques to boost resistance to diseases and pests as well. "But with biotic stresses, it is more complicated because you're defending the plant against pathogens or insects that are evolving," says Dobermann.

Getting durable resistance to insects often requires several genes with different properties, continual improvement, and wise farming practices, as illustrated by the fight against the brown planthopper. The tiny insect sucks the sap from rice stalks and often infects the plant with viruses. Infestation can be deadly. In the 1970s, the planthopper was brought to heel through integrated pest management—which encourages the use of natural predators—and the development of resistant varieties.

But in just 10 years, planthoppers developed an ability to attack resistant plants as well as resistance to a widely used pesticide. Annual losses in China are estimated to run 2.77 million tons and in Vietnam about 700,000 tons, says Kong Luen Heong, an IRRI entomologist. The root problem is overuse of pesticides, which kill off the planthopper's natural predators. "This is a problem of unsustainable practices," Heong says. Breeding resistant varieties might help, he says, but to be effective, new varieties must be integrated with changes in farming practices. IRRI is planning a pest-management demonstration project in China in 2009 that minimizes pesticide use.

Researchers have cultivars that are resistant to other stresses—including drought, cold, and iron toxicity—in the R&D pipeline. Teams are also working on genetically modified (GM) varieties. Public antipathy, particularly in Asia, has kept GM rice confined to labs. A variety modified to produce pro–vitamin A could force governments to come to terms with transgenic crops (*Science*, 25 April, p. 468). IRRI now has the so-called golden rice in a field trial, and trials in farm fields in Bangladesh could start in about 2 years, Dobermann says. But he thinks it will take at least a decade for GM rice to have a significant impact on production.

Another factor slowing work on new varieties is the structure of the rice market. Private companies conduct a lot of research on crops such as maize and soybeans because there is a thriving seed business. Rice farmers, on the other hand, retain part of each season's crop as seed for the next crop, so there is a smaller seed business and advances depend heavily on public-sector efforts. Pardey says little public spending in advanced countries goes to increasing grain productivity; instead, it is spent mostly on fruits and vegetables and environmental concerns. Contributions to organizations like IRRI have waned: IRRI's budget has eroded from a peak of $44.4 million in 1993 to $27.9 million in 2006. And few developing countries, aside from China and India, have been ramping up spending as quickly as they need to, Pardey says. As a result, over the past 10 years maize yields have risen by nearly 1.8% per year while growth in rice yields has slipped below 1% annually and is virtually nil across Asia, Dobermann says.

Closing the gap

Reducing losses to stresses can only partly ameliorate a crisis. Varieties tolerant to submergence, drought, and salinity are useful in environments that account for about 25% of global rice production. "If we want to do something in terms of food security," says Dobermann, "we need to invest much more in improving varieties" for the 75% of rice grown in favorable environments.

SOWING THE SEEDS OF EXPERTISE

Getting his feet wet. Caleb Dresser, an undergrad at Cornell University, learned to till a rice paddy during the IRRI course.

Money alone won't reinvigorate agricultural R&D; fresh talent is needed, too. An innovative training program at the International Rice Research Institute (IRRI) in Los Baños, Philippines, aims to hone the skills of established rice researchers and entice young scientists into the field.

The 3-week course, supported by the U.S. National Science Foundation and the U.K.'s Gatsby Charitable Foundation, was held for the second time in late May and early June. Participants got hands-on experience in how rice is sown, cultivated, and harvested. They also heard about the latest progress in research and plant breeding and discussed practical problems such as fertilizer management with scientists and farmers.

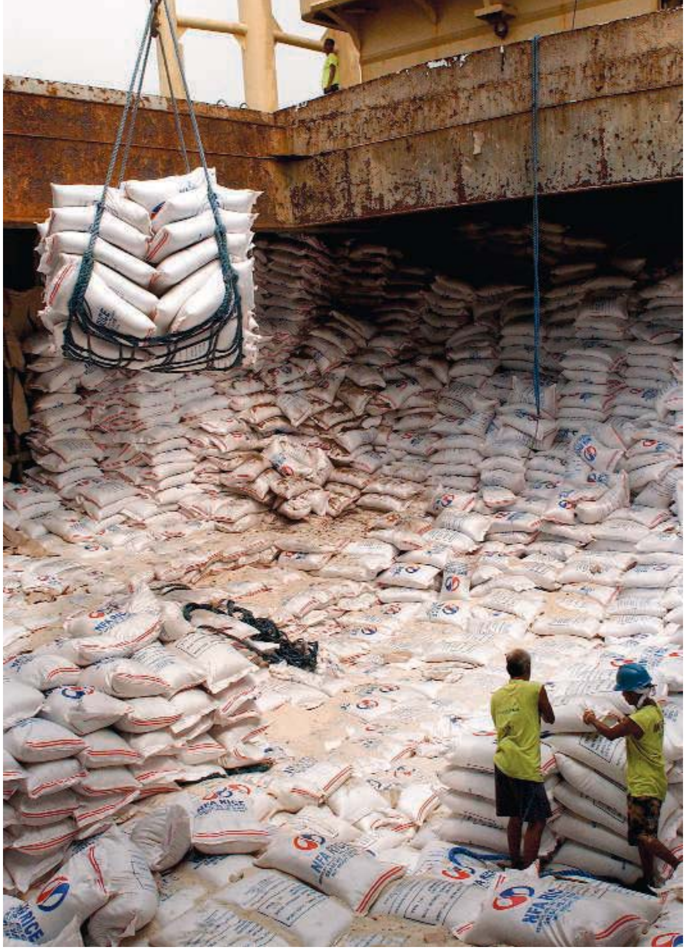
The course, developed by IRRI and Susan McCouch, a rice geneticist at Cornell University, attempts to put rice in a social, economic, and cultural context. "There are students of molecular biology for whom rice is a series of A's and G's and T's and C's," McCouch says, referring to DNA's four nucleotides. Meanwhile, researchers and extension workers from rice-growing countries in Asia and Africa often have little exposure to advanced lab techniques and few contacts.

Roughly half of the 29 participants came from Europe or the United States, and half from Asia or Africa. The "very multicultural and interdisciplinary" mix gave participants a taste of international collaboration, says Margaret Mangan, who starts work on a Ph.D. in agroecology at the University of Minnesota, St. Paul, this fall.

Scientists from rice-growing countries took away methods for improving practices at home. Abubakary Kijoji, a technical assistant with the Eastern and Central Africa Rice Research Network in Dar es Salaam, Tanzania, learned a new irrigation technique in which paddies are watered rather than flooded. "That is something we can apply [to] the challenge of water shortages," he says.

Mangan, who admits that she never saw a rice plant before coming to the Philippines, says the course "helped me remember why I got into agriculture." While concentrating on basic research for her Ph.D., she intends to keep the big picture in mind in hopes of making connections between "the very technical side and the very practical side of agriculture." A few bumper harvests of such scientists might turn today's rice crisis into tomorrow's surplus. **–D.N.**

CREDIT: HARRIS TUMAWIS/IRRI



Precious cargo. A dozen countries have restricted rice exports to protect domestic consumers, pushing export prices to record levels.

CREDIT: RAYMOND PANALIGAN/IRRI

Recent improvements in potential rice yields have been incremental in part because breeders have already picked the low-hanging fruit. In a sign of the challenges ahead, Qifa Zhang, a rice geneticist at Huazhong Agricultural University in Wuhan, China, identified a gene on chromosome 7 that plays a key role in boosting yield potential. He found, however, that most modern cultivars already carry the gene. Understanding how it works might lead to yield gains, says Zhang, whose findings appeared in *Nature Genetics* last May. "But we'll have to be creative in deciding how to make use of it."

Higher yields could come from greater reliance on hybrid rice. Hybrids of genetically diverse plants benefit from heterosis, or hybrid vigor, which produces yields up to 20% greater than inbred varieties. China pioneered the use of hybrid rice in the 1970s and now plants it on 16 million hectares, or 57% of its total rice area. Last year, hybrid rice accounted for about 65% of China's 186 million ton rice production, according to Longping Yuan, director-general of the China National Hybrid Rice R&D Center and a professor at Hunan Agricultural University in Changsha. The average yield of hybrids is 7.1 to 7.2 tons per hectare versus 5.8 to 5.9 tons per hectare for inbred varieties.

But several factors have limited the spread of hybrid rice. Yuan's hybrids are indica varieties suited for the tropics. His team has not yet produced an effective japonica hybrid for temperate regions. In addition, Yuan admits, the hybrid rice he introduced in 1976 "was just so-so" in taste and quality. It was promoted by a central government anxious to feed its people, he says. His center is striving to improve the rice's taste.

Because of quality concerns, breeders in other countries have been slow to adapt hybrids to local conditions. Hybrid rice also requires a change in farming culture and infrastructure. The practice of retaining part of a crop as seed works for inbred varieties that are self-pollinating. But the yield benefit of heterosis is seen only in first-generation crosses. This means new hybrid seed must be purchased for each crop.

The drawbacks have limited hybrid rice to about 4 million hectares outside China. But Dobermann foresees that total rising to as much as 20 million hectares in a decade as varieties improve.

One alternative—looking to wild and exotic strains—promises to boost yields of inbred varieties. For decades, breeders have worked with a limited number of rice varieties chosen for observable traits, says Susan McCouch, a rice geneticist at Cornell University. Wild and exotic varieties were ignored, she says, because they yield less rice than modern cultivars and thus were not obvious sources of beneficial genes.

In the 1990s, McCouch and Cornell colleague Steven Tanksley crossed wild and exotic rice varieties with modern cultivars and then used molecular linkage maps to identify genes in offspring that increased yield. They almost always found some yield-boosting genes from the wild parent, McCouch says. They then added targeted genes from the wild parent to modern cultivars. This strategy appears to have an effect similar to heterosis, but the desired trait is fixed and boosts yields in later generations.

Now about a dozen groups around the world are using wild rice genes in this way to improve local varieties. Sang-Nag Ahn, a rice breeder at Chungnam National University in Daejeon, South Korea, and his colleagues crossed four elite Korean rice cultivars with wild species. Some offspring yielded 10% to 20% more grain than the parents, says Ahn. The most promising lines are in field trials; he expects to release the first of these crosses to farmers in 3 to 5 years.

A more ambitious plan is to convert rice from a C3 to a C4 plant that's better at bulking up on carbon. C3 plants—the majority of species, including wheat, barley, and potatoes—use the enzyme RuBisCO to turn carbon dioxide into a three-carbon compound that is fixed into the plant's biomass. Less common C4 plants, such as maize and sugar cane, have an additional enzyme, PEP carboxylase, which produces a four-carbon compound that RuBisCO fixes more efficiently. C4 plants, which probably evolved from C3 plants millions of years ago, are 50% more efficient at turning sunlight into biomass. John Sheehy, an IRRI plant physiologist, says that a C4 rice plant could boast 50% greater yield while requiring less water and fertilizer (*Science*, 28 July 2006, p. 423).

Sheehy and colleagues have screened wild relatives of rice and found some evidence of the close vein spacing in leaves, the large numbers of photosynthesizing chloroplasts, and the CO_2-absorption characteristics that are typical of C4 plants. "They are not C4 plants but are closer to C4 than normal C3 plants," Sheehy says. He predicts it could take several years to prove that rice can be transformed into a C4 plant and a decade or more to produce a prototype. That's just the kind of long-term, high-payoff research that governments should be funding, says Pardey.

A meta-analysis of hundreds of studies that Pardey's group is preparing for publication shows "a pretty well-established relationship" between R&D and increasing yields. They also found that the peak effect of a discovery comes 20 to 25 years after the research was initiated. Conversely, sagging growth in agricultural productivity is the direct result of limited increases in R&D funding since the late 1970s, Pardey says. Reversing the trend requires "a decadal response," he says, "not a political cycle response." The rice crisis that caught the world off-guard may take many years to resolve.

–DENNIS NORMILE

GENETICS

Simple Sleepers

Classic genetic model organisms—fruit flies, zebrafish, and roundworms—are popular newcomers in sleep research laboratories, although debate continues about how much their dozing relates to human slumber

Joan Hendricks thought she had killed her charges. She had been sitting under a dim red light in a basement for hours, tapping on vials of fruit flies to keep the insects active. Eventually, the flies rolled around seemingly lifeless; her tapping didn't rouse them. But a couple of hours later, Hendricks, a sleep researcher who is now the dean of the University of Pennsylvania veterinary school, realized her flies were simply sacked out. "They were just so sleepy," Hendricks says. "They were basically dead on their little fruit fly feet."

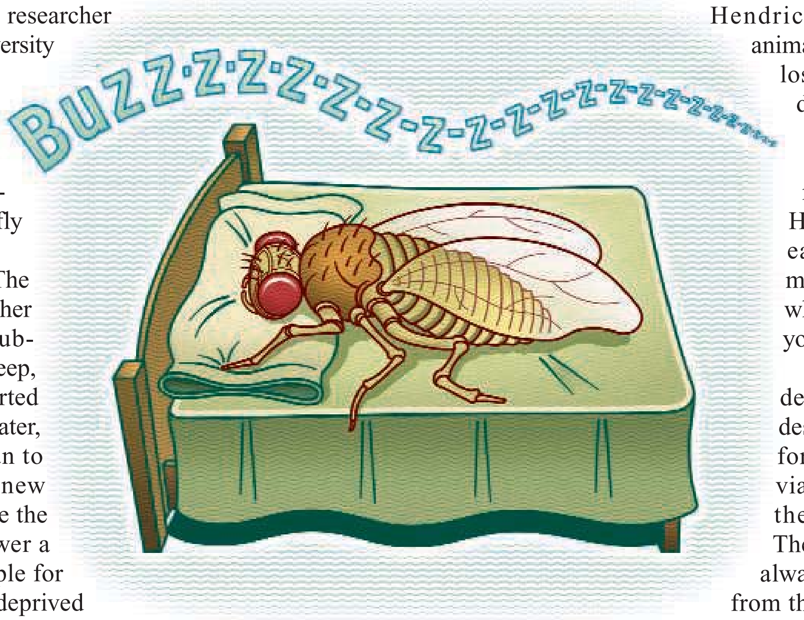


ILLUSTRATION: PETER HOEY

That was in the late 1990s. The experiments by Hendricks and her colleagues led to the first published description of fruit fly sleep, in 2000. A second group reported similar findings a few months later, and the drowsy insects began to usher sleep research into a new molecular age. Scientists hope the new approach will help answer a question that has baffled people for centuries: Why sleep? Sleep-deprived humans feel awful and perform poorly; rats deteriorate and die if they're kept awake for barely more than 2 weeks. But no one knows the reason.

Now, arguing that most organisms slumber much as humans do, a growing number of sleep researchers are welcoming fruit flies, zebrafish, and roundworms—classic simple animal models—into their labs. Already, these creatures, with their easy-to-study genomes and simple nervous systems, have yielded new evidence for how sleep maintains the brain and metabolism. They've also revealed genes that regulate sleep—including a fly gene called *sleepless*, described on page 372.

Not everyone is convinced that fish, fly, and worm sleep will shed light on human slumber, or that sleep even has a common function across the animal kingdom, but some promising parallels have convinced many researchers that they're on the right track. "I'm a true believer," says Chiara Cirelli, a neuroscientist at the University of Wisconsin (UW), Madison, who, like Hendricks, was among the first to study sleep in the fruit fly *Drosophila melanogaster*. "The more we look at them, the more they look very similar to mammals."

Even some human sleep researchers share her optimism. Psychiatrist Eric Nofzinger of the University of Pittsburgh School of Medicine in Pennsylvania, who studies sleep disorders using brain imaging, is eager for the genetic insights that only the simpler organisms can provide. "There is a lot of promise and possibility," he says.

Fatigued flies

There's a reason simple animals such as flies and worms escaped the attention of sleep researchers for so long. When birds and mammals sleep, their brains generate characteristic electrical patterns that denote deep sleep and dreaming. Since discovering this phenomenon in 1953 using electroencephalogram recordings of human brains, scientists have incorporated EEG patterns into the definition of sleep. But the simpler brains of flies, worms, and even reptiles don't produce those patterns, and no one was certain these animals even sleep.

So by the mid-1990s, when new molecular and genetic techniques pioneered in fruit flies and worms were illuminating everything from memory formation to embryonic development, sleep researchers were still stuck with model organisms such as cats, rats, and dogs. But some researchers, such as Allan Pack of the University of Pennsylvania School of Medicine, suspected that fruit flies did snooze, and they hoped that studying flies would similarly illuminate the genetics of sleep and its disorders.

To confirm that flies doze, Pack and his colleagues, including Hendricks, resurrected and refined older behavioral criteria that had been superseded by the EEG: A sleeping animal should be still and difficult to rouse. It should assume a habitual posture or protected location. And most important, Hendricks says, sleep-deprived animals should try to make up for lost slumber, just like people do. This would indicate that sleep isn't just a time-killer but a basic need. "It's regulated like hunger," Hendricks says. "If you don't eat for a long time, you eat more. If you can't sleep for a while, the need builds up and you sleep more."

That's where her sleep-deprived flies came in: They desperately needed to make up for lost sleep after a night of vial-tapping. The insects met the other sleep criteria too. They snoozed mainly at night, always a few millimeters away from their food, and it took more-vigorous tapping to rouse sleeping insects than alert ones, Hendricks and her colleagues reported in *Neuron* in 2000. Coincidentally, Paul Shaw, as a postdoc with neurobiologist Giulio Tononi of the Neurosciences Institute in San Diego, California, was getting similar results, and that group published their work a few months later (*Science*, 10 March 2000, p. 1834).

To see if the molecular underpinnings of sleep were conserved, both teams gave their flies food laced with caffeine at one-quarter to five times the concentration of a cup of coffee. The ones that consumed the highest dose dozed only half as much as caffeine-free controls, and some even died. Other compounds that affect human sleep similarly influenced flies: Amphetamines kept them awake, and antihistamines made them fall asleep. Even the pattern of rest over the flies' roughly 2-month life span was reminiscent of that of mammals: The youngest flies slept the most and elderly flies the least.

The two fly-sleep papers started a trend. Researchers published behavioral and

Published by AAAS

pharmacological evidence of slumbering zebrafish (*Danio rerio*) in 2002 and dozing roundworms (*Caenorhabditis elegans*) in 2008. As for flies, there are excellent genetic tools for both fish and worms, and the two animals are well-suited to studies relating sleep to nervous system structure and maintenance.

The worm, however, has an odd sleep schedule. From the time it hatches, *C. elegans* takes just a few days to mature. Rather than sleep daily like flies and other animals, the growing worm takes a 2-hour nap (a state called lethargus) every 7 to 12 hours at each of four developmental transitions. During these periods, the worm builds a new cuticle, restructures body parts, and, finally, reaches sexual maturity. From then on, at least in the lab, the worm never sleeps again. David Raizen, a neurologist who studies *C. elegans* sleep at the University of Pennsylvania Medical School, says the contrast to mammalian sleep is actually a good thing: Lethargus is so different that anything the two have in common is probably important to sleep's universal function. "The trick," Raizen says, "is to look at similarities."

Promising parallels

Indeed, deep homologies have begun to emerge among the cell-signaling systems that promote sleep and wakefulness in different species. Earlier this decade, Hendricks and a team led by neurobiologist Amita Sehgal at the University of Pennsylvania Medical School found that mutant flies with excessive signaling from a transcription factor called CREB slept up to 50% more than normal flies. Some flies with a crippled CREB system slept less than half the usual amount and had an abnormally long sleep rebound after deprivation. Two years later, a research group that included Pack found that mutant mice lacking CREB also slept less than controls. Similar congruence has emerged for epidermal growth factor (EGF) signaling, which promotes sleep in worms, flies, rabbits, and hamsters.

Still, Pack points out, molecules such as CREB and EGF "are involved in lots of [biological] processes," from storing memories to governing cell fate during development. They couldn't represent a dedicated sleep mechanism, he contends. Rather, Pack says, sleep must emerge from the combined action of these signals on neural circuits in the brain.

Pack predicts that flies, worms, and fish could help sort out how brain regions communicate to produce sleep. He cites the neurotransmitter dopamine, which acts only on a subset of neurons in flies and mammals—and in both, it promotes wakefulness. Similarly, in both groups of animals, GABA promotes sleep. What's more, the genetic tools available for these simple animals allow researchers to map which brain regions respond to specific sleep-regulating molecules. In flies, CREB promotes wakefulness by acting on the mushroom bodies, a part of the insect brain in charge of learning and memory. And the EGF pathway affects sleep through two regions of the fly brain and exactly one *C. elegans* neuron.

Yet it's not obvious how such findings will translate to humans. "The brain neurochemistry and architecture are fundamentally different," warns Emmanuel Mignot, a sleep researcher at Stanford University in Palo Alto, California, who pioneered the study of narcolepsy in dogs. He points out that, unlike dopamine and GABA, some critical neurotransmitters involved in mammalian sleep are entirely absent in flies and worms. Hypocretin, for example, has been linked to narcolepsy in mice, dogs, and people, but flies and worms don't even make it. Given such differences, Mignot doesn't expect flies and worms to reveal much about mammalian sleep at the level of the whole brain. Rather, he says, the utility of these animals is to uncover functions and mechanisms of sleep in individual cells.

Insomniac. Flies with a mutation in the sleepless gene (*right*) are active even while normal flies sleep, as seen in this 10-minute composite photo.

CREDITS (TOP TO BOTTOM): CHIARA CIRELLI; KYUNGHEE KOH

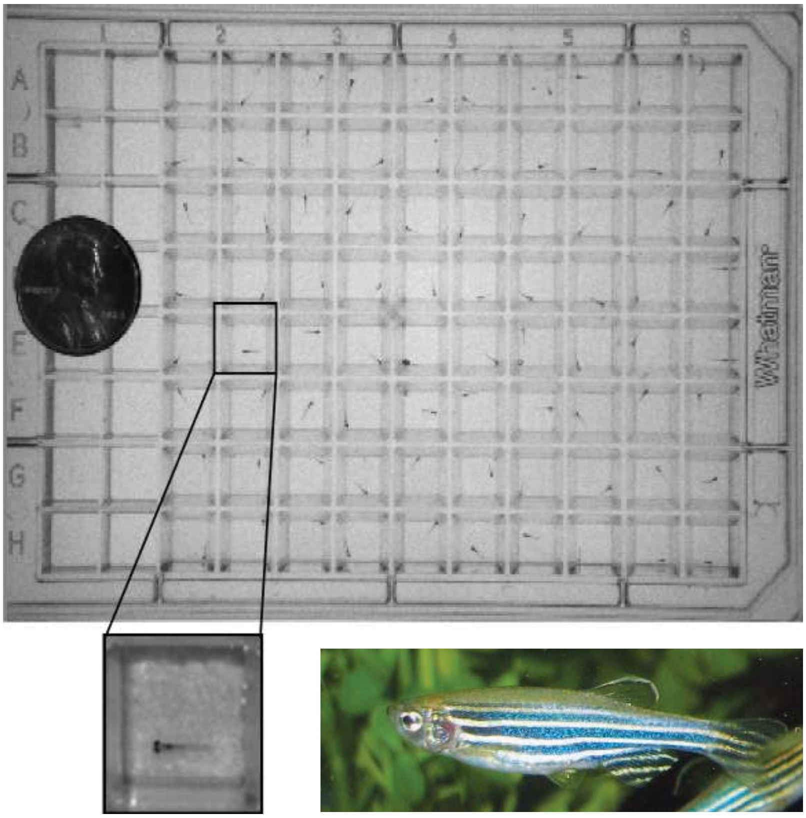
Rough night. This robot periodically jostles vials of fruit flies, keeping the insects awake for sleep-deprivation studies.

Finding a function

One way to get at the basic cellular purpose of sleep is to compare which genes and proteins are active only during sleeping or waking. In mice, rats, sparrows, and flies, numerous genes involved in protein synthesis and cholesterol metabolism work mainly during sleep. An accumulation of such research, including their own mouse studies, led Pack and colleagues to propose in 2007 that a key function for sleep is to give the body time and energy to rebuild molecules that are used up during waking. The *C. elegans* nap cycle squares with this idea, Raizen says. During lethargus, the worms synthesize a new skin-like cuticle and double the cell nuclei in their intestines, even though the cells themselves don't divide. "Those are two intensely biosynthetic events," he says.

Tired insects have helped suggest that the nervous system may need sleep for a related reason. When comparing genes in rats and flies that are active during waking and turned off during sleep, Cirelli and Tononi, both now at UW Madison, noticed that several were involved in building and strengthening synapses, connections among neurons in the brain that are a result of learning. If all the new synapses accumulated day after day, the brain would soon run out of space and energy, says Cirelli. (Already, the brain accounts for one-fifth of human metabolism.) She suggests that during sleep, the synapses are trimmed so that only the most robust connections remain. "You don't lose the memories, but … you wake up with a leaner brain," Cirelli says.

That hypothesis remains to be tested, but there is a general sense that sleep has something to do with taking the nervous system offline for maintenance. Sleep-deprived humans and rats perform poorly in mental tasks from college exams to mazes—and flies are looking quite similar. Shaw, now at Washington University in St. Louis, Missouri, and his colleagues have done experiments, not yet published, demonstrating that sleep-deprived flies make about 50% more mistakes than well-rested flies in a learning test.



Like us. Zebrafish larvae fit in a 96-well plate for large-scale screening of how drugs affect their sleep. Their brain chemistry and anatomy have much in common with other vertebrates, including humans.

CREDIT: D. A. PROBER *ET AL., THE JOURNAL OF NEUROSCIENCE* 26, 51 (20 DECEMBER 2006)

Looking forward

Like many in the sleep research community, Sehgal suspects there's some big idea about the function of sleep that has been missed. "We're looking for fundamental advances," she says. The best way to find such a new idea, she says, is to screen for abnormal sleep among flies, worms, and fish whose DNA has been mutated at random positions. Such screens avoid any preconceived notions about what sleep does. "You just … identify the genes that correspond to those mutations," Sehgal says, and ultimately, "maybe the function will fall out of those things."

In 2005, Cirelli and her team reported in *Nature* the first dividend of such a random genetic screen. Her team observed 9000 different mutant fly lines for altered sleep need. One of the most extreme lines slept only 4 to 5 hours per day, compared with the normal 9 to 15 hours—and remained alert even after sleep deprivation. These restless flies, which tended to die earlier than normal, turned out to have a mutation in a gene called *Shaker*, which encodes a protein channel that controls the flow of potassium across cell membranes. Functional Shaker channels help neurons return to baseline after firing; defects in them increase neuronal excitability and had been previously linked to epilepsy but not to sleep.

Unlike flies, which have one version of the *Shaker* gene, mice have more than a dozen *Shaker*-like genes. Cirelli's team rolled the dice, picked one, and introduced a mutation into it. The resulting mutant rodents spent 21% more time awake than normal mice, the team reported in *BMC Biology* in 2007.

In the new *Science* paper, Sehgal's team reports another extravagantly short-sleeping fly strain that underlines the importance of the Shaker channel and neuronal excitability for sleep. Created for a genetic screening effort, these flies had a mutation in a previously undescribed gene and as a consequence slept less than an hour per day. The team then found a second line of flies with a different mutation in the same gene. Although these flies had a normal amount of spontaneous daily sleep, they had almost no rebound following deprivation. Sehgal's team dubbed the new gene *sleepless* and ultimately showed that it codes for a small protein in the brain, a protein that may regulate expression of the Shaker channel: *sleepless* mutants had barely detectable levels of the Shaker channel protein.

Mignot says the *Shaker* and *sleepless* mutants together point to a previously unsuspected importance of neuronal excitability in sleep. "It's an idea that has emerged from [these mutant flies]," he says. Still, it's not clear that neuronal excitability varies across the natural sleep-wake cycle. "I'm worried," he says, "that … it could be a pathological way to induce waking."

To get the vertebrate side of the sleep story, both Mignot and Harvard University neurobiologist Alexander Schier have begun their own randomized genetic screens using zebrafish—and Schier's team has also tested thousands of drugs for their effects on zebra-fish sleep in a study yet to be published. Mignot predicts rapid progress as flies, worms, and fish reveal new sleep genes and begin to cross-pollinate with human research. "In the next 5 years, there will be an avalanche of knowledge," he says.

Others are less certain. Jerome Siegel, who studies regulation of REM sleep in mammals at the University of California, Los Angeles, worries that his colleagues, in their rush to embrace the new animal models, are overinterpreting similarities among species. "We have to appreciate that there is a tremendous diversity of sleep, even within mammals," he says. Indeed, giraffes may snooze only a few hours per day, while some bats sack out 20 hours at a stretch. And dolphins rest only half their brains at any one time. "Looking for a brain function for sleep doesn't make a lot of sense," Siegel says. Rather, he argues, sleep could just be a way to save energy and to stay out of trouble once other needs are met.

Mignot disagrees, insisting that there is a 90% consensus that sleep has some restorative function—even if that function is yet to be revealed. And at the very least, researchers studying flies, worms, and fish can now dream of solving that grand mystery.

–ELSA YOUNGSTEADT

Former *Science* intern Elsa Youngsteadt is a freelance writer in Raleigh, North Carolina.

MEETINGBRIEFS>>

ACOUSTICS '08 | 29 JUNE–4 JULY | PARIS, FRANCE

Sonic history. Stanford archaeologist John Rick (*right*) is creating an acoustic map of prehistoric underground chambers in Peru.

Sound Science Maps Venetian Canals and Peruvian Ruins

Think of archaeologists out in the field and you probably picture them carrying shovels and sample bags—perhaps also a bullwhip if you're an Indiana Jones fan. But the newest generation of researchers may be as likely to wield sensitive microphones and recorders, according to several sessions in Paris devoted to archaeological acoustics.

Some are harnessing sound waves to reveal the invisible. Fantina Madricardo, a geologist at the Institute of Marine Science in Venice, Italy, presented a dramatically different map of her famous city's lagoon. Using a newly developed shallow-water sonar system, she and colleagues charted subtle differences in sediment density, tracing the contours of canals and structures buried for millennia beneath the shifting water. The sonar map led archaeologists to dig in places "where they did not think to look," Madricardo says. A previously undiscovered Roman brick embankment is now being unearthed. Because river deltas have been a favorite site for human occupation going back to Neolithic times, the method could reveal many more submerged artifacts elsewhere.

Online
sciencemag.org
S Listen to roaring tigers, snorting polar bears, a cracking mountain of ice, and more from Acoustics '08.

Other archaeological studies are turning to sound not as a tool but as an artifact itself. "Ancient soundscapes have been largely ignored by archaeology," says David Lubman, a veteran acoustical scientist who is now an industry consultant based in Westminster, California. One reason is that "sound is ephemeral," so reconstructing what an environment sounded like hundreds or even thousands of years ago is a daunting task. Then, he says, "comes the question of intentionality—how do you prove that people were aware of particular acoustic phenomena?"

For example, for about a decade, Lubman has argued that the Maya Temple of Kukulkan step pyramid in Chichen Itza, Mexico, was designed with sound in mind. In Paris, he convinced scientists in the audience that the temple's peculiar shape makes it reflect the sound of clapping back as a chirp that closely matches the call of the quetzal, a bird revered by the Maya. But few of his listeners agreed that the effect was deliberately built into the temple. "I do not believe it," says Jorge Cruz, an acoustics Ph.D. student at the Mexican National Polytechnic Institute in Mexico City.

The question of intentionality may also arise from an ongoing study described by Miriam Kolar and Patty Huang, acoustics Ph.D. students at Stanford University in Palo Alto, California. As part of a team led by Stanford archaeologist John Rick, they are mapping the acoustic environment within the 3000-year-old labyrinthine galleries of Chavín de Huántar in the Peruvian highlands. Conch-shell trumpets previously discovered within the galleries indicate that the underground chambers were used for music and ritual ceremonies. The new work shows that the layout of the galleries creates reverberations that make it impossible to pinpoint sound sources. Rick hypothesizes that a ruling priest class exploited the disorienting acoustic effect to instill awe and fear. The study "is persuasive because it builds on evidence of performance in the space," says Christopher Scarre, an archaeologist at Durham University in the U.K.

In collaboration with Stanford composer John Chowning, the team plans to use its three-dimensional acoustic map of the galleries to create a virtual simulation of the rituals that took place at Chavín de Huántar. "In a couple of years, the galleries are going to be reinforced to prevent collapse, which will change their acoustics," says Kolar. "So this is really the last chance to preserve this ancient soundscape."

Ultrasound Uses in Medicine Heat Up

Hospitals have used ultrasound for decades to see fetuses and kidney stones without breaking a patient's skin. Now doctors' sonic toolkits are about to expand, as new ultrasound-based technologies are poised to probe the inner structure of bones and treat otherwise incurable cancers.

X-ray photography "only gives you the average density of the bone," says Pascal Laugier, a medical researcher at Université Pierre et Marie Curie in Paris. "To determine the quality of bone, you need to see its inner structure, and only ultrasound can reveal this." That information is particularly crucial in aging populations of the industrialized countries, where the burden of treating bone fractures from osteoporosis and falls has ballooned. As a bonus, ultrasonic devices require none of the radioisotopes or heavy shielding of x-ray machines.

The bone biophysicists gathered in Paris compared notes on the many problems with interpreting the ultrasonic sounds that propagate through the body's hardest material. A team led by Victor Humphrey, a physicist at the University of Southampton, U.K., presented an acoustic study of how bones heal after a break. Bones typically form a thick "callus" of tissue around fractures, cement the gap, and then reabsorb the callus. A bone's stage of healing can be quickly assessed, Humphrey reports, by transmitting ultrasound through the fracture—

CREDIT: COURTESY OF JOHN RICK/STANFORD

Published by AAAS

Snapshots From the Meeting >>

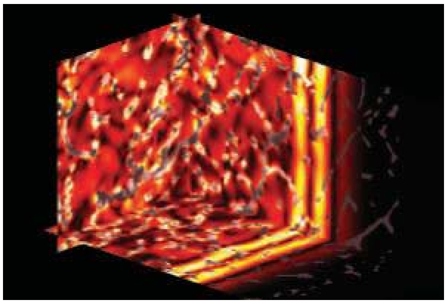
Sleepy talking. Slumped posture and bloodshot eyes are giveaways of extreme lack of sleep, but speech betrays fatigue, too. A team led by Suzanne Boyce, a linguist at the University of Cincinnati in Ohio, recorded people giving directions after a good night's rest and after 34 to 58 hours of sleep deprivation. By analyzing phonetic features of each person's speech—pauses and the loss of syllables—a computer spotted a pattern associated with drowsiness. "This work is very exciting," says Sarah Hawkins, a linguist at the University of Cambridge, U.K. "It promises to … not only help with practical applications such as detecting when machine operators like airline pilots are tired but will also give us greater insight into … how speech is produced."

How polar bears and tigers hear. The results of the first hearing tests of large carnivores were welcomed by conservationists eager to know the frequency ranges at which these animals can perceive human noise. A team led by Anne Bowles, a biologist at the Hubbs–Sea World Research Institute in San Diego, California, trained polar bears to respond to tones in order to receive a snack. The bears had a hearing range similar to that of humans, between 125 and 20,000 hertz. Meanwhile, tiger hearing was tested at Henry Doorly Zoo in Omaha, Nebraska. A team led by Edward Walsh, a physiologist at nearby Boys Town National Research Hospital in Omaha, analyzed the spectrum of tiger roars and also used electrodes to monitor brain activity in anesthetized tigers while playing a range of sounds. The results, says Walsh, support theories that tigers communicate with each other by infrasound, sound of lower frequency than most mammals perceive. "Confrontational roars" contain infrasonic energy, and other tigers can hear it, Walsh says. But such sounds are absent from the "territorial roars" that tigers use to maintain their vast domains. **–J.B.**

at least for long bones such as those in the arm. But one of the most medically important bones still defies ultrasound analysis. "We have not figured out the hip," says Laugier. "It has such complex geometry and diverse structure."

Cancer researchers discussed high-intensity focused ultrasound (HIFU) therapy, which focuses ultrasound beams on a single spot inside the body, such as the center of a tumor. The mechanical energy of the acoustic waves converts into heat, and the tissue dies in a sizzle. HIFU is already used to treat cancer in easily accessible tissue such as in the prostate and uterus, but the brain has so far been off-limits because the skull makes focusing the beams nearly impossible. The solution, says Mathias Fink, a physicist and medical researcher at the University of Paris, is "time-reversal acoustics," a strategy that uses echo patterns as a guide for focusing waves through a barrier. Fink has used the technique to successfully target brain tumors in animals.

As the hype for HIFU grows—it's widely used to treat cancer in China and is being evaluated in the United States and Europe—some scientists at the meeting urged caution. HIFU often generates transient microscopic bubbles that can boost the temperature of the surrounding tissue to potentially dangerous levels. "No simple way exists … to calculate temperature rise and therapeutic dose in tissue," says Peter Kaczkowski of the University of Washington, Seattle. The ultrasound engineer calls for more basic research on HIFU and better "regulatory oversight."



Bone probe. Passing ultrasonic waves through bone, as in this numerical simulation of a section of femur, reveals its inner structure.

CREDITS (TOP TO BOTTOM): PHOTOS.COM; E. BOSSY AND P. LAUGIER, *IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL*, 55, 6 (© JUNE 2008 IEEE)

Listening to Distant Ice Crack

Using satellites, scientists have kept a wary eye on the crumbling Antarctic ice shelf, tracking the movement of titanic chunks that break free. Now, using acoustic instruments designed to detect nuclear explosions, they are putting an ear to the ice as well.

Big breaks in the ice shelf over the past 2 decades have been dramatic, but it remains unclear whether they are due to global warming. That's partly because most cracks and breaks are too small to be seen from space. Getting a statistical handle on those smaller events, said polar scientists at the meeting, will help determine whether the rate of ice-shelf degradation stays within natural bounds or steadily increases.

The sound of an ice mountain cracking quickly dissipates through air and land, but it can propagate through water for thousands of kilometers. So Alexander Gavrilov and Binghui Li, marine acousticians at Curtin University of Technology in Perth, Australia, took advantage of a Cold War legacy: three hydrophone arrays in the Indian Ocean that listen for nuclear explosions as part of the Comprehensive Nuclear Test Ban Treaty. Each 2-kilometer-wide array reveals the direction of sounds; by triangulating data from the three stations, researchers can trace the location of the sound sources.

The first result is that the system works. "Antarctica is the major source of low-frequency noise in the Southern Ocean and southern parts of the Indian Ocean," Gavrilov reports. To confirm that these sounds originated from distant ice cracking, he and Li compared a year of the sounds with recordings from a hydrophone they installed on the Antarctic sea floor.

Next, the scientists focused on the dozen or so daily "cracking and breaking events" from the ice shelf that could be detected in the Indian Ocean data. Over the past 7 years, they found seasonal variation in the sounds but no significant increase—or decrease—over time. This is "good news," says Gavrilov, because the data set can be used as a baseline for monitoring the ice in coming years.

"Seismologists a decade ago would have never dreamed that these kinds of signals would be broadcast from the ice masses into the far fields of the world's oceans," says Douglas MacAyeal, a geophysicist at the University of Chicago in Illinois. But there may be limits to the interpretation of these acoustic signals. So far, says Ursula Schauer, a geophysicist at the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven, Germany, scientists cannot use the sounds to calculate the volume of ice breaking off the shelves. "But this might be possible in the future with more sophisticated signal processing," she says. The study is a "major innovation," adds Terence Hughes, a glaciologist at the University of Maine, Orono, but more ears in the water are needed. "It should be employed all around Antarctica, not just in the Indian Ocean sector." **–JOHN BOHANNON**

Published by AAAS

LETTERS

edited by Jennifer Sills

The Cost of Conservation

C. KREMEN *ET AL.* ("ALIGNING CONSERVATION PRIORITIES ACROSS TAXA IN MADAGASCAR WITH high-resolution planning tools," Reports, 11 April, p. 222) proposed a systematic plan for acquiring new protected areas in Madagascar, using extensive new species richness data, but their analysis did not consider the costs of acting in different regions. Costs vary substantially; omitting this important facet of conservation planning can lead to poor biodiversity outcomes.

Conservation agencies are increasingly incorporating realistic costs to optimize future actions, with the help of conservation software (*1*, *2*). Analyses have shown that the including costs can considerably increase the efficiency of conservation plans, by up to a factor of 10, compared with plans that use area as a proxy (*3*–*5*). Estimated land costs in Madagascar (*6*) vary by up to four orders of magnitude (between USD $0.60 and $1785 per hectare), and some areas identified as priorities by Kremen *et al.* are in Madagascar's most expensive regions. The costs of the priority areas identified mirror the overall distribution of costs in Madagascar, whereas a more efficient solution would favor low-cost areas. Given that large areas of Madagascar have relatively low opportunity costs, much more biodiversity could have been protected with the same investment.

Madagascar's wildlife. Preserving diversity in Madagascar depends in part on pursuing the most effective conservation strategy.

In the developing world, such as Madagascar, conservation decisions that do not include the opportunity costs to stakeholders are unlikely to effectively protect biodiversity (*7*–*9*). High-cost sites are usually in demand for other purposes, and targeting these sites for conservation will cause conflict with people who depend on this land. If planners do not attempt to avoid conflict with local stakeholders by including their values throughout the planning process, then reserves will be prone to failure. "Paper parks" are a reality in many developing countries (*8*), including Madagascar (*9*), where disenfranchised local communities ignore park boundaries. Local groups are also more likely to suffer from injudicious protected area placement if costs are not included. The resulting expulsions lead to loss of livelihood and cultural degradation (*10*), while robbing the conservation movement of effective political allies (*8*). Before Kremen *et al.*'s methods are used to guide conservation actions, the varying costs of conservation must be incorporated.

MICHAEL BODE,[1]* JAMES WATSON,[2] TAKUYA IWAMURA,[2] HUGH P. POSSINGHAM[2]

[1]Applied Environmental Decision Analysis Group, School of Botany, The University of Melbourne, Melbourne, VIC 3010, Australia. [2]Applied Environmental Decision Analysis Group, School of Integrative Biology, The University of Queensland, Brisbane, QLD 4072, Australia.

*To whom correspondence should be addressed. Email: mbode@unimelb.edu.au

References

1. H. P. Possingham, I. Ball, S. Andelman, in *Quantitative Methods for Conservation Biology*, S. Ferson, M. Burgman, Eds. (Springer-Verlag, New York, 2000).
2. A. Moilanen *et al.*, *Proc. R. Soc. London Ser. B* **272**, 1885 (2005).
3. A. Ando, J. Camm, S. Polasky, A. Solow, *Science* **279,** 2126 (1998).
4. R. Naidoo, A. Balmford, P. J. Ferraro, S. Polasky, T. H. Ricketts, M. Rouget, *Trends Ecol. Evol.* **21,** 681 (2007).
5. S. Polasky *et al.*, *Land Econ.* **77**, 68 (2001).
6. R. Naidoo, T. Iwamura, *Biol. Conserv.* **140,** 40 (2007).
7. D. Brockington, *Fortress Conservation: The Preservation of Mkomazi Game Reserve, Tanzania* (James Currey, Oxford, 2002).
8. J. B. Alcorn, *Conserv. Biol.* **7**, 424 (1993).
9. N. Seddon *et al.*, *Oryx* **34**, 287 (2000).
10. W. M. Adams *et al.*, *Science* **306**, 1146 (2004).

Conservation with Caveats

C. KREMEN *ET AL.* ("ALIGNING CONSERVATION priorities across taxa in Madagascar with high-resolution planning tools," Reports, 11 April, p. 222) identified the optimal sites for expansion of Madagascar's land area under protection, using advanced conservation planning techniques at an unprecedented level of detail for six taxonomic groups. However, the exhaustive study had caveats.

Conservation planning analyses that incorporate biodiversity value and economic costs, unlike that of Kremen *et al*., show that limited budgets can achieve substantially larger biological gains than plans that ignore costs (*1*–*3*).

This may seem trivial in Madagascar's exceptional case, as the targets and timeline for conservation are set. But imagine if Madagascar's President Ravalomanana were able to conserve more than 10% land area and preserve more biodiversity, for the same cost, if he focused on getting the most "bang for his buck" (*4*).

In addition, climate change is already causing shifts in species ranges (*5*), which will likely change the future battlegrounds for conservation (*6*). This underscores the desperate need to incorporate climate change into conservation planning.

Finally, as the authors rightly point out, the analysis would benefit from the inclusion of other taxa. In particular, we wonder whether conservation priorities would change if a well-known taxon like birds had been included.

The analysis is an advancement to secure Madagascar's biodiversity at a crucial and opportune time. Still, it should be disconcerting for conservationists that it nonethe-

CREDIT: JUPITER IMAGES

Origin of vertebrate vocals
347

Past glacial extents in the Southern Hemisphere
348

less has caveats. These are key areas for future research.

BERNARD W. T. COETZEE

Centre for Invasion Biology, Department of Zoology and Entomology, University of Pretoria, South Africa, and WWF Office, Suva, Fiji. E-mail: bwtcoetzee@zoology.up.ac.za

References

1. R. Naidoo *et al.*, *Trends Ecol. Evol*. **21**, 681 (2006).
2. K. A. Wilson *et al.*, *PLoS Biol.* **5**, 1850 (2007).
3. K. A. Wilson *et al.*, *Nature* **440**, 337 (2006).
4. H. P. Possingham, K. A. Wilson, *Nature* **436**, 919 (2005).
5. C. Parmesan, G. Yohe, *Nature* **421**, 27 (2003).
6. T. M. Lee, W. Jetz, *Proc. R. Soc. London Ser. B,* 10.1098/rspb.2007.1732 (2008).

Response

BODE *ET AL.* NOTE THAT WE SHOULD INCLUDE cost data in our analysis. Our analysis optimized conditions for biodiversity to produce a pure "biodiversity benchmark" against which plans incorporating other elements can be compared. Only with such a benchmark can one quantify what would be lost when political or socioeconomic constraints are accommodated. That said, we recognize that the costs of conservation are spatially heterogeneous and that high-resolution cost surfaces should be used when available to maximize biodiversity benefits under cost constraints (*1*). Using data from a global-scale analysis of opportunity costs (*2*), Bode *et al.* imply that the existing and proposed protected areas are equivalent in cost to a random sample across Madagascar, rather than to a solution that minimizes cost. However, the global-scale opportunity cost data they used cannot provide a good metric for assessing the cost consequences of the existing plus proposed areas in Madagascar for several reasons: (i) The global-scale data appear to be grossly inaccurate for some regions of the country. For example, the southern and western areas shown as being most productive for crops and livestock (see figure 1A in Bode *et al.*'s ref. 1) are in fact Madagascar's most arid, drought-ridden regions, occupied by its poorest people, and subject to frequent famines. In contrast, some major regions for producing Madagascar's staple crop, rice, are shown incorrectly to have low opportunity cost (e.g., rice-producing regions around Lake Alaotra). (ii) The resolution of the global-scale map of opportunity costs is 100 times coarser than our analysis. It is inappropriate to use global-scale data, whether economic or biological, to develop or to evaluate subregional conservation plans (*3*). (iii) This global opportunity cost layer is based solely on agricultural and livestock production, omitting several high-value potential land uses in Madagascar (mining, timber production, and ecotourism). While Bode *et al.* hint that our efforts might have been better spent

Published by AAAS

generating a spatially heterogeneous cost rather than biodiversity surface, we suggest that it would be rash to generalize from the relatively few studies [e.g., (*4*, *5*)], mostly global scale, that have discussed the relative utility of cost versus biodiversity data for achieving biodiversity outcomes.

Bode *et al.* add that conservation plans will not succeed without making local people central to the strategy. In Madagascar and elsewhere, implementing real conservation plans involves multiple iterations of discussion between policy-makers and both natural and social scientists (*6*). Members of our research team have been actively involved with the multi-institutional body governing such discussions (Système d'Aires Protegées) since its inception. At the national scale, our results [e.g., figure 2B in (*7*)] are being integrated with other, previously generated biodiversity priority areas for Madagascar, expert knowledge of current habitat condition, predictions of deforestation threats and ecosystem service benefits, local stakeholder interests, climate change refugia (from an expert workshop held in Madagascar in January 2008), and consideration of mining and ecotourism interests. After this national synthesis, more detailed, bottom-up planning at the local to regional scale will utilize both additional layers and stakeholder input, before protected areas are finally delimited, zoned, and gazetted [for an example of this process, see (*8*)]. We agree with Bode *et al.* that biodiversity will not ultimately be conserved without taking local and national political, social, and economic concerns into account, and reiterate the value of a quantified biodiversity "benchmark" in multisectoral decision-making for conservation.

Coetzee adds that we omitted the effects of climate change and bird data. We incorporated basic design elements for climate change by maximizing the proportions of species' ranges included and by prioritizing landscape connectivity. Our analysis prioritized small-ranged species (*7*), which are more vulnerable to climate change (*9*). Many of these small-ranged species occur around mountain tops, which are ultimately expected to become climatic refuges (*9*–*12*). Their inclusion in the network builds in some resilience to climate change.

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Although we were unable to access bird data suitable for our modeling procedure, Important Bird Areas and IUCN Extent of Occurrence data for threatened birds were heavily used in planning the prior expansion of reserves (2002 to 2006).

Optimization techniques contribute to efficient conservation planning, but it is important to understand the limitations of such products. Solutions, while "optimized," are probably never "optimal."

CLAIRE KREMEN,[1]* ALISON CAMERON,[1] ANDRIAMANDIMBISOA RAZAFIMPAHANANA,[2] ATTE MOILANEN,[3] CHRIS D. THOMAS,[4] HENK BEENTJE,[5] JOHN DRANSFIELD,[5] BRIAN L. FISHER,[6] FRANK GLAW,[7] TATJANA C. GOOD,[8] GRADY J. HARPER,[9] ROBERT J. HIJMANS,[10] DAVID C. LEES,[11] EDWARD LOUIS JR.,[12] RONALD A. NUSSBAUM,[13] STEVEN J. PHILLIPS,[14] CHRISTOPHER J. RAXWORTHY,[15] GEORGE E. SCHATZ,[16] MIGUEL VENCES,[17] DAVID R. VIEITES,[18] PATRICIA C. WRIGHT,[19] MICHELLE L. ZJHRA[8]

[1]Department of Environmental Sciences, Policy, and Management, University of California, Berkeley, CA 94720–3114, USA. [2]REBIOMA, Wildlife Conservation Society, BP 8500, Antananarivo 101, Madagascar. [3]Metapopulation Research Group, Department of Biological and Environmental Sciences, University of Helsinki, Post Office Box 65, Viikinkaari 1, FI-00014 Finland. [4]Department of Biology (Area 18), University of York, Post Office Box 373, York YO10 5YW, UK. [5]Royal Botanical Gardens, Richmond TW9 3AB, Surrey, UK. [6]Department of Entomology, California Academy of Sciences, San Francisco, CA 94103, USA. [7]Zoologische Staatssammlung München, Münchhausenstrasse 21, 81247 München, Germany. [8]Department of Biology, Georgia Southern University, Statesboro, GA 30460, USA. [9]Center for Applied Biodiversity Science, Conservation International, 2011 Crystal Drive, Suite 500, Arlington, VA 22202, USA. [10]International Rice Research Institute, Los Baños, Philippines. [11]Department of Entomology, Natural History Museum, London SW7 5BD, UK. [12]Center for Conservation and Research, Henry Doorly Zoo, Omaha, NE 68107, USA. [13]Museum of Zoology, University of Michigan, Ann Arbor, MI 48109–1079, USA. [14]AT&T Labs-Research, 180 Park Avenue, Florham Park, NJ 07932, USA. [15]American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024–5192, USA. [16]Missouri Botanical Garden, Post Office Box 299, St. Louis, MO 63166–0299, USA. [17]Zoological Institute, Technical University of Braunschweig, 38106 Braunschweig, Germany. [18]Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720–3160, USA. [19]Department of Anthropology, State University of New York, Stony Brook, NY 11794, USA.

*To whom correspondence should be addressed. E-mail: ckremen@nature.berkeley.edu

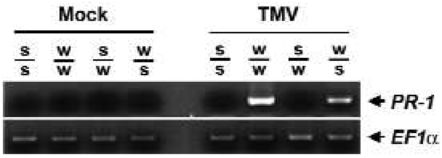
References

1. R. Naidoo *et al.*, *Trends Ecol. Evol.* **21**, 681 (2006).
2. R. Naidoo, T. Iwamura, *Biol. Conserv.* **140**, 40 (2007).
3. M. Roguet, *Biol. Conserv.* **112**, 217 (2003).
4. M. Bode *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6498 (2008).
5. P. J. Ferraro, *J. Pol. Anal. Manage.* **22**, 27 (2003).
6. A. T. Knight, R. M. Cowling, B. R. Campbell, *Conserv. Biol.* **20**, 408 (2006).
7. C. Kremen *et al.*, *Science* **320**, 222 (2008).
8. C. Kremen *et al.*, *Conserv. Biol.* **13**, 1055 (1999).
9. W. Thuiller, S. Lavorel, M. B. Araujo, M. T. Sykes, I. C. Prentice, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8245 (2005).
10. T. L. Root *et al.*, *Nature* **421**, 57 (2003).
11. C. Parmesan, G. Yohe, *Nature* **421**, 37 (2003).
12. L. Hannah *et al.*, *Front. Ecol. Env.* **5**, 131 (2007).

CORRECTIONS AND CLARIFICATIONS

Reports: "ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis" by K. Ishikawa *et al.* (2 May, p. 661). The 25 February 2008 submission date was incorrect. The correct submission date was 10 September 2007.

Reports: "Methyl salicylate is a critical mobile signal for plant systemic acquired resistance" by S.-W. Park *et al.* (5 October 2007, p. 113). Two lanes in Fig. 1E may have been duplicated. They are S/S and S/W under TMV.



Therefore, the results were independently confirmed in a double-blind experiment. The new data are presented here and confirm the results originally presented. EF1α was used as an internal control. Semiquantitative RT-PCR, rather than RNA blot analysis, was used to quantify PR-1 transcript levels.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea"

Thijs J. G. Ettema and Siv G. E. Andersson

Berg *et al.* (Reports, 14 December 2007, p. 1782) reported the discovery of a novel autotrophic carbon dioxide–fixation pathway in Archaea and implicated a substantial role of this pathway in global carbon cycling based on sequence analysis of Global Ocean Sampling data. We question the validity of the latter claim.

Full text at www.sciencemag.org/cgi/content/full/321/5887/342b

RESPONSE TO COMMENT ON "A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea"

Ivan A. Berg, Daniel Kockelkorn, Wolfgang Buckel, Georg Fuchs

We proposed that the 3-hydroxypropionate/4-hydroxybutyrate cycle might be important in global carbon cycling based on the abundance of related autotrophic Crenarchaea in the ocean and the high number of gene sequences for a key enzyme of the cycle. Here, we counter the specific criticisms raised by Ettema and Andersson.

Full text at www.sciencemag.org/cgi/content/full/321/5887/342c

Comment on "A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea"

Thijs J. G. Ettema* and Siv G. E. Andersson

Berg *et al*. (Reports, 14 December 2007, p. 1782) reported the discovery of an autotrophic carbon dioxide–fixation pathway in Archaea and implicated a substantial role of this pathway in global carbon cycling based on sequence analysis of Global Ocean Sampling data. We question the validity of the latter claim.

Berg *et al*. (*1*) reported the discovery of a CO_2-fixation pathway in autotrophic members of the archaeal order Sulfolobales, as well as in members of the Cenarchaeales and Archaeoglobales. This pathway, which the authors refer to as the 3-hydroxypropionate/4-hydroxybutyrate pathway, is shown to comprise a cycle of 16 enzymes, one of which is the proposed key enzyme 4-hydroxybutyryl-CoA dehydratase (4HCD). Based on a comparison of abundances of 4HCD and RuBisCo sequences in the Global Ocean Sampling (GOS) data (*2*), Berg *et al.* (*1*) predicted the abundance of the newly discovered CO_2-fixation pathway in the ocean surface waters to be of the same order of magnitude as the Calvin-Bassham-Benson cycle, which is known to be of global importance for carbon cycling. In addition, the authors proposed the existence of an abundant group of mesophilic autotrophic Crenarchaea in the ocean surface waters that uses the proposed 3-hydroxypropionate/4-hydroxybutyrate pathway for CO_2 assimilation. We raise several concerns about the validity of these claims.

First, the abundance of 4HCD homologs does not necessarily reflect the abundance of the 3-hydroxypropionate/4-hydroxybutyrate pathway, because this enzyme is also known to participate in 4-aminobutyrate fermentation in a few strict anaerobic bacteria (*3*). Moreover, homologs of 4HCD are implicated in unrelated metabolic processes, including phenylacetate catabolism (*4*) and pyoverdine chromophore biosynthesis (*5*).

Second, solely comparing the "number of identified…gene sequences" (*1*) in metagenomics data does not provide an accurate estimate of the relative abundances of the enzyme or the species, because these numbers depend on the criteria used for homolog identification, gene copy numbers, presence of paralogs, rates of gene sequence evolution, cloning efficiencies, and lengths of the individual genes, or even the choice of filter-fraction size for sample collection. For example, the relatively large cell sizes of cosmopolitan cyanobacteria such as *Synechococcus* and *Prochlorococcus* has likely resulted in a considerable underrepresentation of such sequences in the GOS data set (*6*), and hence of Calvin-Bassham-Benson cycle sequences. Thus, attempts to infer the abundance of the novel CO_2-fixation pathway relative to the (cyanobacterial) Calvin-Bassham-Benson cycle based on the number of sequence reads with gross similarity to 4HCD and RuBisCo, respectively (*1*), should be considered with caution.

Third, below the euphotic zone (>150 m), pelagic crenarchaeota are known to make up a large fraction of total marine picoplankton (*7*) and to perform pivotal roles in marine nitrogen and carbon cycles (*8*). Based on an analysis of the GOS ocean surface metagenome data, Berg *et al.* alluded to the existence of a "group of

---

Department of Molecular Evolution, Evolutionary Biology Centre, Uppsala University, Uppsala, S-752 36 Sweden.

*To whom correspondence should be addressed. E-mail: thijs.ettema@ebc.uu.se

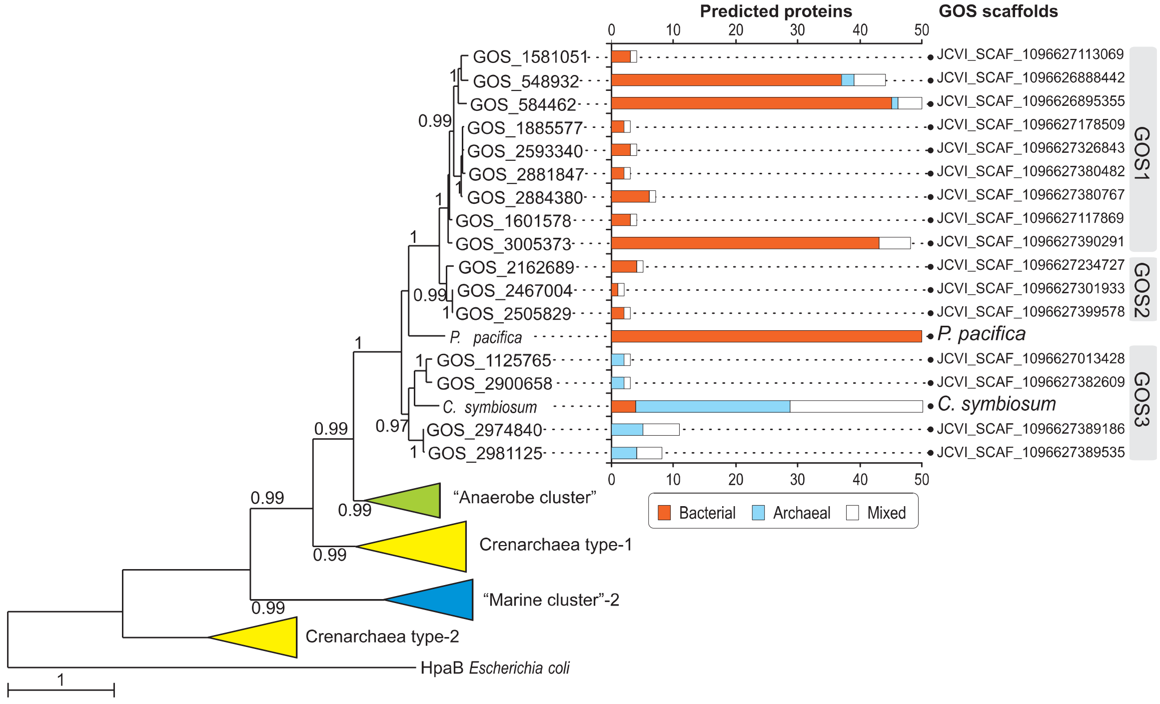
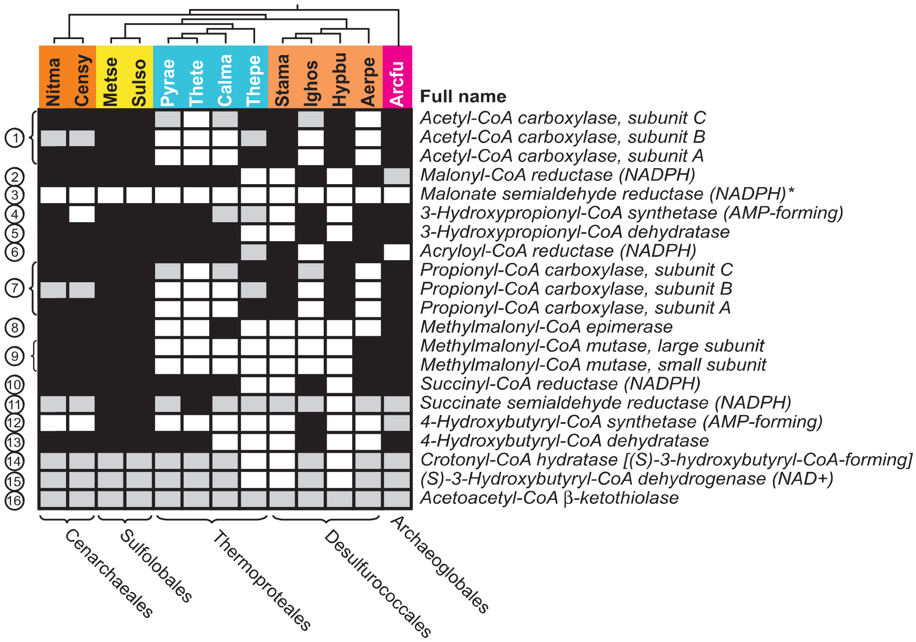


Fig. 1. Phylogenetic relationships of 4HCD protein sequences and a kingdom analysis (*9*) of proteins encoded by the genes flanking the 4HCD sequences in the GOS scaffolds. The GOS1 and GOS2 clades are dominated by bacterial genes (also see fig. S1), indicating that these 4HCD gene products are unlikely to function in the proposed archaeal 3-hydroxypropionate/4-hydroxybutyrate pathway for CO_2 assimilation. As a control, the results for the same BLASTP-search-based method are depicted for 50 random proteins from *C. symbiosum* and *P. pacifica*, respectively (excluding self-hits). The phylogenetic tree of the 4HCD proteins is based on the same sequence alignment as in Berg *et al.* (*1*, *9*). The scale bar represents a difference of 1 substitution per site, and the numbers at the nodes indicate the resampling estimated log-likelihood (RELL) support values. Only RELL values above 0.95 are shown. Major groups that gave a similar overall topology as found by Berg *et al*. [figure 3 in (*1*)] have been collapsed and depicted as triangles.



	Nitma	Censy	Metse	Sulso	Pyrae	Thete	Calma	Thepe	Stama	Ighos	Hypbu	Aerpe	Arcfu	Full name
①	■	■	■	■	▨	☐	▨	■	■	▨	■	☐	■	*Acetyl-CoA carboxylase, subunit C*
	▨	▨	■	■	☐	☐	☐	▨	■	☐	■	☐	■	*Acetyl-CoA carboxylase, subunit B*
	■	■	■	■	☐	☐	☐	■	■	☐	■	☐	■	*Acetyl-CoA carboxylase, subunit A*
②	■	■	■	■	■	■	■	☐	☐	■	☐	■	▨	*Malonyl-CoA reductase (NADPH)*
③	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	*Malonate semialdehyde reductase (NADPH)**
④	■	☐	■	■	■	■	▨	▨	☐	■	☐	■	■	*3-Hydroxypropionyl-CoA synthetase (AMP-forming)*
⑤	■	■	■	■	■	■	■	☐	☐	■	☐	■	■	*3-Hydroxypropionyl-CoA dehydratase*
⑥	■	■	■	■	■	■	■	▨	■	☐	■	■	☐	*Acryloyl-CoA reductase (NADPH)*
⑦	■	■	■	■	▨	☐	▨	■	■	▨	■	☐	■	*Propionyl-CoA carboxylase, subunit C*
	▨	▨	■	■	☐	☐	☐	▨	■	☐	■	☐	■	*Propionyl-CoA carboxylase, subunit B*
	■	■	■	■	☐	☐	☐	■	■	☐	■	☐	■	*Propionyl-CoA carboxylase, subunit A*
⑧	■	■	■	■	☐	☐	■	☐	☐	☐	☐	☐	■	*Methylmalonyl-CoA epimerase*
⑨	■	■	■	■	☐	☐	☐	☐	☐	☐	☐	■	■	*Methylmalonyl-CoA mutase, large subunit*
	■	■	■	■	☐	☐	☐	☐	☐	☐	☐	■	■	*Methylmalonyl-CoA mutase, small subunit*
⑩	■	■	■	■	■	■	■	☐	☐	■	☐	■	■	*Succinyl-CoA reductase (NADPH)*
⑪	▨	▨	■	■	▨	■	▨	▨	▨	▨	☐	▨	▨	*Succinate semialdehyde reductase (NADPH)*
⑫	☐	☐	■	■	☐	☐	☐	☐	☐	■	☐	☐	▨	*4-Hydroxybutyryl-CoA synthetase (AMP-forming)*
⑬	■	■	■	■	■	■	☐	☐	☐	■	☐	☐	■	*4-Hydroxybutyryl-CoA dehydratase*
⑭	▨	▨	▨	▨	▨	▨	▨	☐	☐	▨	☐	▨	▨	*Crotonyl-CoA hydratase [(S)-3-hydroxybutyryl-CoA-forming]*
⑮	▨	▨	▨	▨	▨	▨	▨	☐	☐	▨	☐	▨	▨	*(S)-3-Hydroxybutyryl-CoA dehydrogenase (NAD+)*
⑯	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	*Acetoacetyl-CoA β-ketothiolase*

Fig. 2. Distribution of the 16 enzymes that make up the 3-hydroxypropionate/4-hydroxybutyrate pathway across archaeal genomes (*9*). Gene presence was inferred from the archaeal cluster of orthologous groups of proteins (*12*) (in black shading) or BLASTP searches (in gray shading). The patchy distribution pattern across species hints at the existence of variants of the 3-hydroxypropionate pathway for CO_2 assimilation (such as in *C. symbiosum* and *N. maritimus*) or at enzymes that connect to or operate in other pathways (*13*). Species abbreviations: Nitma, *Nitrosopumilus maritimus*; Censy, *Cenarchaeum symbiosum*; Metse, *Metallospaera sedula*; Sulso, *Sulfolobus solfataricus* P2; Pyrae, *Pyrobaculum aerophilum*; Thete, *Thermoproteus tenax*; Calma, *Caldivirga maquilingensis* IC-167; Thepe, *Thermofilum pendens* Hrk 5; Stama, *Staphylothermus marinus* F1; Ighos, *Ignicoccus hospitalis*; Hypbu, *Hyperthermus butylicus*; Aerpe, *Aeropyrum pernix*; Arcfu, *Archaeoglobus fulgidus*. Enzymes: 1, acetyl-CoA carboxylase; 2, malonyl-CoA reductase (NADPH); 3, malonate semialdehyde reductase (NADPH); 4, 3-hydroxypropionyl-CoA synthetase (AMP-forming); 5, 3-hydroxypropionyl-CoA dehydratase; 6, acryloyl-CoA reductase (NADPH); 7, propionyl-CoA carboxylase; 8, methylmalonyl-CoA epimerase; 9, methylmalonyl-CoA mutase; 10, succinyl-CoA reductase (NADPH); 11, succinate semialdehyde reductase (NADPH); 12, 4-hydroxybutyryl-CoA synthetase (AMP-forming); 13, 4-hydroxybutyryl-CoA dehydratase; 14, crotonyl-CoA hydratase [(S)-3-hydroxybutyryl-CoA-forming]; 15, (S)-3-hydroxybutyryl-CoA dehydrogenase (NAD$^+$); 16, acetoacetyl-CoA β-ketothiolase.

abundant mesophilic autotrophic crenarchaea" in the ocean surface waters, which they propose use the 3-hydroxypropionate/4-hydroxybutyrate pathway for CO_2 assimilation (*1*). This hypothesis is not supported by data from the GOS expedition (*2*), which indicated that archaeal sequences were less abundant by a factor of ~33 than bacterial sequences (2.7% and 90.8% of total reads, respectively) (*6*). Of the 60 most abundant 16S rRNA ribotypes, *Prochlorococcus* and *Synechococcus* show 171-fold coverage of the 16S rRNA gene as compared to 8-fold coverage for Archaea (*2*). Thus, there is at least a 20-fold difference between the most abundant cyanobacterial and archaeal species in the upper surface waters.

These apparent discrepancies prompted us to reexamine the data that brought Berg *et al.* to their conclusion. We carefully reanalyzed the proposed phylogeny of 4HCD protein sequences (*9*) and, like Berg *et al.* (*1*), observed that the GOS environmental sequences formed a cluster that was more closely related to a clade of anaerobic bacterial species that use 4HCD in fermentation than to the crenarchaea that use this enzyme for CO_2 fixation [figure 3 in (*1*)]. Within the "marine cluster"–1, the GOS1 and GOS2 groups clustered with high support, whereas the proposed 4HCD from the mesophilic crenarchaeaote *Cenarchaeum symbiosum* was found to cluster within the GOS3 group (Fig. 1). The 4HCD homolog from the strictly aerobic, chemoheterotrophic marine myxobacterium *Plesiocystis pacifica* is also a member of the "marine cluster"–1, but its exact placement could not be resolved in either analysis [Fig. 1 and figure 3 in (*1*)].

Because the phylogenetic analysis fails to distinguish between a bacterial versus an archaeal origin of the environmental sequences in "marine cluster"–1, we analyzed the remainder of the (predicted) genes encoded by the scaffolds on which the 4HCD sequences are located. Sequences that reside on the same scaffold originate from the same species, and an analysis of all genes encoded by the scaffold should in principle distinguish bacterial from archaeal sequences. A search for sequence similarity of proteins encoded by the GOS scaffold using the basic local alignment search tool for proteins (BLASTP) (*9*) revealed that the scaffolds associated with GOS1 and GOS2 groups are dominated by bacterial genes and are thus likely of bacterial origin (Fig. 1). The scaffolds associated with the GOS3 group of "marine cluster"–1 contain mostly archaeal genes and might thus be of archaeal origin. Because the 3-hydroxypropionate/4-hydroxybutyrate CO_2 assimilation pathway is thought to be restricted to archaeal species, the most plausible explanation is that the bacterial 4HCD-like sequences that dominate "marine cluster"–1 are most likely involved in another pathway.

Furthermore, 4HCD is sensitive to oxygen exposure due to inactivation of the [4Fe-4S] iron-sulfur clusters that reside in its active center. The enzyme might be sufficiently stable at the thermophilic, and thus low-oxygen pressure, environments in which members of the Sulfolobales reside. However, the oxygen-rich conditions in the ocean surface waters might require additional measures to prevent inactivation. Interestingly, the amino acid residues that constitute the iron-sulfur cluster pocket are not conserved in the "crenarchaea type-2" 4HCD sequences [figure 3 in (*1*)], hinting at a possible adaptation toward oxygen exposure. It remains unclear how the seemingly intact iron-sulfur clusters from "marine cluster"–1 4HCDs resist oxygen inactivation.

Finally, Berg *et al.* predict the 3-hydroxypropionate/4-hydroxybutyrate pathway to be operational in *C. symbiosum* and *Archaeoglobus fulgidus* (*1*). Because we were unable to detect candidate genes for some of the components of the pathway in these species, this conclusion seems premature. For example, candidates for succinyl/malonyl-CoA reductase, succinate-semialdehyde reductase, and the 4-hydroxybutyryl-CoA and 3-hydroxypropionyl-CoA synthetases are missing or cannot be conclusively identified for *C. symbiosum* (Fig. 2). Thus, one cannot rule out that these organisms rely on yet another variant of the 3-hydroxypropionate pathway for CO_2 fixation (*10*). Future experimental studies of the autotrophic pathways of mesophilic crenarchaeaotes, such as *Nitrosopumilus maritimus*, should resolve this issue. The anticipated diversity in archaeal CO_2 assimilation pathways resembles the diversity observed in other archaeal central carbon metabolic pathways (*11*).

References and Notes

1. I. A. Berg, D. Kockelkorn, W. Buckel, G. Fuchs, *Science* **318**, 1782 (2007).
2. D. B. Rusch *et al.*, *PLoS Biol.* **5**, e77 (2007).
3. A. Gerhardt, I. Cinkaya, D. Linder, G. Huisman, W. Buckel, *Arch. Microbiol.* **174**, 189 (2000).
4. M. A. Prieto, J. L. Garcia, *J. Biol. Chem.* **269**, 22823 (1994).
5. A. Stintzi *et al.*, *J. Bacteriol.* **181**, 4118 (1999).
6. S. Yooseph *et al.*, *PLoS Biol.* **5**, e16 (2007).
7. M. B. Karner, E. F. DeLong, D. M. Karl, *Nature* **409**, 507 (2001).
8. A. E. Ingalls *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6442 (2006).
9. Materials and methods are available as supporting material on *Science* Online.
10. S. J. Hallam *et al.*, *PLoS Biol.* **4**, e95 (2006).
11. C. H. Verhees *et al.*, *Biochem. J.* **375**, 231 (2003).
12. K. S. Makarova, A. V. Sorokin, P. S. Novichkov, Y. I. Wolf, E. V. Koonin, *Biol. Direct* **2**, 33 (2007).

13. The identification of a protein with similarity to a given enzyme does not necessarily imply that the enzymatic activity is conserved. For example, the inferred candidate for malonyl/succinyl-CoA reductase identified for *C. symbiosum* and *N. maritimus* might rather correspond to aspartate semialdehyde dehydrogenase [figure S2 in (*1*)]. Likewise, the predicted C subunit of the acetyl-CoA carboxylase in *A. fulgidus* might correspond to pyruvate carboxylase subunit A. Hence, laboratory experiments are needed to conclusively reveal the nature and the components of the CO_2-fixating pathway(s) that (might) operate in these organisms. The gene encoding NADPH-dependent malonate semialdehyde reductase was not identified in the study by Berg *et al*. (*1*), and multiple candidates were proposed for the genes encoding crotonyl-CoA hydratase, NAD^+-dependent 3-hydroxybutyryl-CoA dehydrogenase, and acetoacetyl-CoA β-ketothiolase. For a full overview of the distribution of predicted orthologs across all sequenced archaeal genomes, including protein identifiers, see (*9*) and table S1.
14. T.J.G.E. acknowledges the support of a Rubicon Fellowship by the Netherlands Organisation for Scientific Research and a Marie Curie Intra-European Fellowship by the European Union. S.G.E.A. acknowledges the support of the Swedish Research Council, the Swedish Foundation for Strategic Research, the European Union, the Göran Gustafsson Foundation, and the Knut and Alice Wallenberg Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/321/5887/342b/DC1
Materials and Methods
Fig. S1
Table S1
References

7 April 2008; accepted 20 June 2008
10.1126/science.1158766

MEDICINE

When Sleep Proves Impossible

Michael L. Perlis

Insomniac

by Gayle Greene

University of California
Press, Berkeley, 2008.
517 pp. $29.95.
ISBN 9780520246300.

Insomniac is worlds apart from other books available on the subject of sleeplessness. What's so different? The book bridges several approaches to a topic that is usually considered from a single vantage point. What Gayle Greene (who teaches literature and women's studies at Scripps College, California) has produced is partly a scholarly work, partly a self-help volume, partly a consumer guide to medical services, partly a scathing op-ed, and partly an autobiography. Frankly, the book is a "cranky" (to use the author's own word) and quirky mix of gold and diamonds with pyrite and cubic zirconia. In the following paragraphs, I summarize some of the aspects of the book and sort them into these two categories.

I'll begin with the positive, the gold and diamonds. The author's perspective that insomnia is an orphan "disease" is, without question, correct. At the research level, there is no part of the National Institutes of Health (NIH) that calls the disorder its own. Greene highlights the research situation by pointing out that in 2005 Sanofi-Aventis's $123 million marketing budget for the sedative Ambien was six times what NIH spent on grant-funded research on insomnia. At the clinical level, while insomnia is the majority sleep disorder in terms of population prevalence, most sleep disorder centers do not have staff clinicians who specialize in its treatment. The shortage of specialists is underscored by the fact that there are only 104 clinicians currently credentialed in behavioral sleep medicine (the branch of sleep medicine that has insomnia as a primary focus).

Greene rightly points out that there is a disconnect between the many forms of insomnia (its diverse nosology) and the tendency to study sleeplessness solely in terms of primary (otherwise referred to as psychophysiological) insomnia. There is no doubt that it would be greatly informative to compare and contrast how the primary entity differs from, e.g., paradoxical and idiopathic forms.

The author offers exceptional summaries regarding the consequences of insomnia, the neurobiology of sleep-wake control, and the putative functions of sleep. She peppers her account with interesting speculations about how γ-aminobutyric acid (GABA) deficiency, luteinizing hormone (which causes the ovaries to release the egg) and orexin hypersecretion, and "hyperarousal" of the central nervous system may each contribute to the incidence or the severity of insomnia. She has substantially enriched the text by including insights and reflections of patients with insomnia along with an abundance of thought-provoking quotations from novelists (some minor, some major) to renowned scholars. For example, Greene quotes a character from Anton Chekhov's *The Cherry Orchard*: "If there's any illness for which people offer many remedies, you may be sure that particular illness is incurable"—which sounds about right to me.

Intermixed with, and overlying, these layers of gold and diamonds is a good deal of pyrite and cubic zirconia. Greene's account gives very little credit to the substantial changes in perspective that have occurred over the past 20 years. For example, when occurring with other diseases insomnia was considered only as a symptom; now, rather than seen as "secondary insomnia," such sleeplessness is classified as a disorder, "comorbid insomnia." This distinction makes it much more likely that, in the future, insomnia will be a focus of more research and targeted treatment.

The author also gives short shrift to the theories regarding the etiology and pathophysiology. Perhaps this is because she finds the majority of these models to be too "psychological." Although it may be true that that there has been a tendency to over-"psychologize" the disorder, and it is certainly true that that the neurobiology of insomnia has been inadequately studied, Greene misses the inherent value of the behavioral and cognitive models and the biological implications of the neurocognitive model. Further, her perspective on cognitive behavioral therapy for insomnia (CBT-I) is too simplistic and reductionist—she boils it down to "change your attitude, change your ways." She fails to appreciate that CBT-I aims to normalize sleep-related homeostatic and circadian dysregulation.

In addition, Greene does not provide a fair review of the efficacy of the various approaches to treatment. In the short term, both pharmacotherapy and CBT-I produce comparable change (with about a 50% reduction in the disturbance of sleep continuity). Medical interventions generally produce changes more rapidly, whereas CBT-I produces more durable changes—changes that are stable in the absence of continued treatment and, for some percentage of patients, gains that continue to accrue with time.

Lastly, the book falls short in its considerations of the "future of insomnia research and new prospects for treatment." For example, the author does not adequately assess currently funded experimental research or delineate how such efforts may be informative. Nor does she adequately review and evaluate the therapies that are presently being studied: behavioral approaches such as CBT-I supplemented by newly developed cognitive therapies, CBT-I combined with bright light treatment, and the new intensive sleep retraining protocol; pharmacological treatments including single-isomer or formulations of existing sedatives, orexin antagonism, selective serotonin antagonism, and the daytime application of modafinil (alone or in combination with CBT-I).

In sum, patients will find *Insomniac* readable, engaging, and sympathetic to their plight and cause. Scholars and scientists will find the book to be a rare and thorough view of the phenomenology of insomnia—but also a frustrating mix of facts and factoids, in which too many ideas and perspectives are summarily endorsed or dismissed. Yet while "quirky" and "cranky," Greene's book is also remarkably comprehensive, and her frustration is also our own: after all this work, we still don't have the answers.

10.1126/science.1159339

The reviewer is at the Sleep and Neurophysiology Research Laboratory, University of Rochester Medical Center, 300 Crittenden Boulevard, Rochester, NY 14642–8409, USA. E-mail: michael_perlis@urmc.rochester.edu

CREDIT: BEATA SZPURA/ILLUSTRATION WORKS/CORBIS

PHILOSOPHY OF SCIENCE

Addressing Complexity

Kim Sterelny

Re-Engineering Philosophy for Limited Beings
Piecewise Approximations to Reality

by William C. Wimsatt

Harvard University Press, Cambridge, MA, 2007.
468 pp. $49.95, £32.95, €37.50.
ISBN 9780674015456.

Idealization is an essential tool of science. The phenomena of the natural world cannot be understood in their full complexity all in one go. We have to sneak up on the world via initial simplifications that isolate the most crucial components of natural systems and their most crucial interactions. Thus random mating and nonoverlapping generations are often built into models of evolutionary transitions, even when it is known that the population in question does not satisfy those assumptions. No matter; complications can be added later. Idealization is also a crucial tool in understanding science itself, which is in its own right an extraordinarily puzzling feature of our social and cultural world. What explains its unique success in generating an increasing understanding of the natural world? What is the nature of that success, and in what ways are those successes limited or constrained? To understand both the triumph and limits of science, we need to proceed via ideal models of scientific domains, products, and practitioners. We can add messy complications later. But we need to choose the right models.

In the rich and impressive collection of essays gathered as *Re-Engineering Philosophy for Limited Beings*, Bill Wimsatt argues that philosophy of science, in its standard forms, has chosen the wrong models: the wrong models of scientists, of their products, and of their explanatory targets. He holds that we are limited beings attempting, with surprising success, to gain predictive and explanatory traction on a complex and intertwined world. We do so with partial, multiple, and plastic approximations. Wimsatt (a philosopher and evolutionary biologist at the University of Chicago) argues that much metascience idealizes away from the interconnectivity of scientific domains, from the cognitive limits of scientists, and from the use of multiple, partial approximations.

So, for example, classic philosophy of science takes the ideal product of scientific practice to be theories. In the best cases, scientific theories have large scope; they have few primitive conceptual elements; they are mathematically precise; and they are complete accounts of their domain of application. Distinct theories of gravitation, or of atomic structure, are rivals: we cannot accept and use more than one. So although on this classical conception of science we sometimes use multiple models of a single system, we do so as a temporary expedient, on the way to a complete theory. Theories are the result of scientific reasoning, and classical philosophy of science treats scientists as ideally rational agents—even though, notoriously, different brands of philosophy of science have very different accounts of ideal rationality. Bayesian approaches are currently in vogue, but many practicing scientists will be more familiar with the Popperian model of the scientist as an ideal refutation engine. Theories, in turn, are theories of domains. On this conception, the world is structured hierarchically into levels of organization. For example, cells are organized structures of macromolecules, themselves built out of biochemical structures that have atomic constituents, which themselves have several layers of subatomic organization. In principle, there are clean reductive relations that show how macrophenomena (like cellular behavior) are composed from, and explained by, interactions between the constituents.

Giuseppe Arcimboldo's *Vertumnus (Emperor Rudolf II)*, 1590.

It is common ground between Wimsatt and his targets that these ideas about science are idealizations, perhaps even extreme ones. But Wimsatt argues that they are unhelpful idealizations. For they idealize away from what we most need to explain: the cognitive success of limited beings. Treating science as ideally rational is like a developmental biologist using preformationism to model development: the subject matter of the discipline has been idealized away. Moreover, Wimsatt argues at length that our cognitive limits are not just a bug. We have developed strategies—imperfect, fallible, but often successful strategies—of learning from our errors. Fallible, context-dependent heuristics are central to the success of science, so a good model of science must focus on understanding the construction and use of heuristics. This task is especially challenging, for these heuristics work despite the fact that the target of science is, often, not a discrete, well-behaved level of organization but a "causal thicket." Causal arrows point down, not just up. Furthermore, activity at a level—say, the behavior of a cell—is not just influenced by adjacent levels.

The world is messy. We are fallible and bounded. Yet science progresses with great reliability. Wimsatt's conception of science is organized around these three facts. Like science itself, his account is partial and incomplete, an approximation organized around the idea of a heuristic. Many questions are left open, and much could be challenged. But Wimsatt is among the most creative, original, and empirically informed philosophers of our day. These essays clearly demonstrate his imagination, his mastery of many diverse literatures, and his eye for the big question. About half have been published before, but mostly in obscure locations. So it is a fine idea to bring these thoughts together, and for the essays to be organized by a detailed connecting narrative. Few essay collections are integrated and systematic: *Re-Engineering Philosophy for Limited Beings* is an important exception.

10.1126/science.1156895

The reviewer is in the Philosophy Program, Research School of Social Sciences, Australian National University, and the Department of Philosophy, Victoria University of Wellington, Kelburn Parade, Wellington 6140, New Zealand. E-mail: kim.sterelny.vuw.ac.nz

CREDIT: SLOTT, SKOKLOSTER, SWEDEN/PHOTO BY ERICH LESSING/ART RESOURCE, NY

ECOLOGY

Assisted Colonization and Rapid Climate Change

Moving species outside their historic ranges may mitigate loss of biodiversity in the face of global climate change.

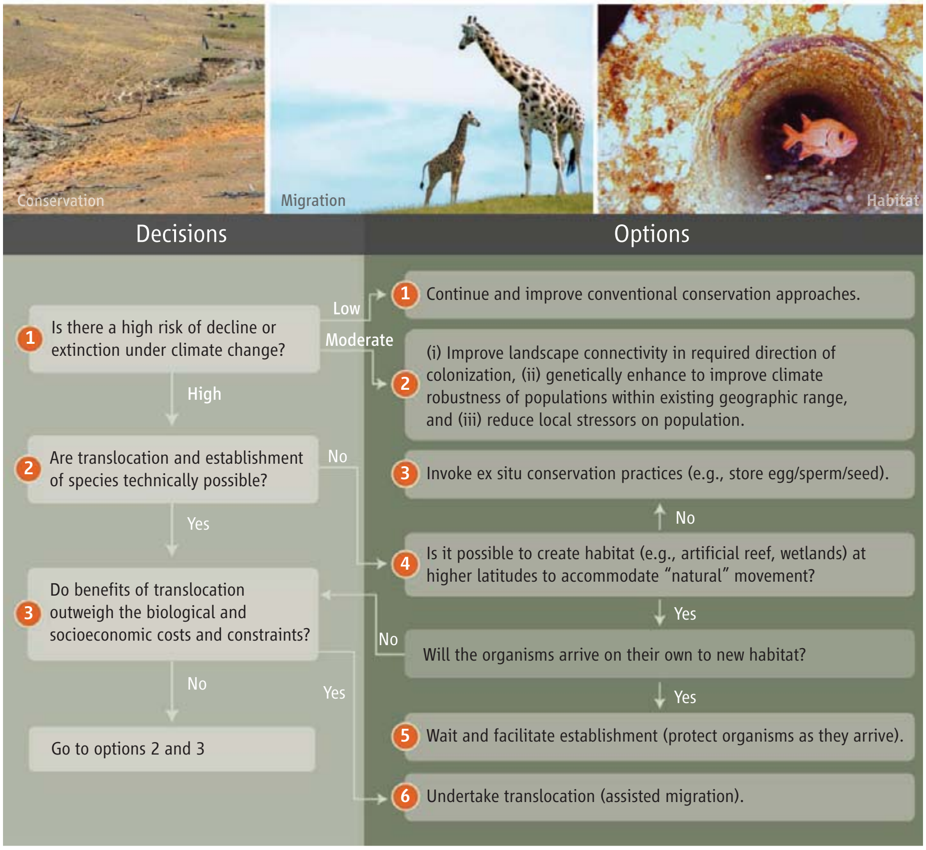
O. Hoegh-Guldberg,[1]* L. Hughes,[2] S. McIntyre,[3] D. B. Lindenmayer,[4] C. Parmesan,[5]
H. P. Possingham,[6] C. D. Thomas[7]

Rapid climatic change has already caused changes to the distributions of many plants and animals, leading to severe range contractions and the extinction of some species (*1*, *2*). The geographic ranges of many species are moving toward the poles or to higher altitudes in response to shifts in the habitats to which these species have adapted over relatively longer periods (*1*–*4*). It already appears that some species are unable to disperse or adapt fast enough to keep up with the high rates of climate change (*5*, *6*). These organisms face increased extinction risk, and, as a result, whole ecosystems, such as cloud forests and coral reefs, may cease to function in their current form (*7*–*9*).

Current conservation practices may not be enough to avert species losses in the face of mid- to upper-level climate projections (>3°C) (*10*), because the extensive clearing and destruction of natural habitats by humans disrupts processes that underpin species dispersal and establishment. Therefore, resource managers and policy-makers must contemplate moving species to sites where they do not currently occur or have not been known to occur in recent history. This strategy flies in the face of conventional conservation approaches. The world is littered with examples where moving species beyond their current range into natural and agricultural landscapes has had negative impacts. Understandably, notions of deliberately moving species are regarded with suspicion. Our contrary view is that an increased understanding of the habitat requirements and distributions of some species allows us to identify low-risk situations where the benefits of such "assisted colonization'" can be realized and adverse outcomes minimized.

Previous discussions of conservation responses to climate change have considered assisted colonization as an option (*11*, *12*), but have stopped short of providing a risk assessment and management framework for how to proceed. Such frameworks could assist in identifying circumstances that require moderate action, such as enhancement of conventional conservation measures, or those that require more extreme action, such as assisted colonization. These frameworks need to be robust to a range of uncertain futures (*13*). Uncertainties arise in climate projections and in how species and ecosystems will respond. Hence, calculation of the lower and upper bounds for the probability and cost of a range of possible outcomes may be the best strategy.

With this in mind, we developed a decision framework that can be used to outline potential actions under a suite of possible future climate scenarios (see figure, below). Determining whether a species faces significant risk of decline or extinction under climate change requires an in-depth knowledge of the underlying species' biology as well as the biological, physical, and chemical changes occurring within its environment. The risk of extinction for many widespread, generalist species found across a range of habitats may be low. In this case, the option of moving such species outside their present



1 Is there a high risk of decline or extinction under climate change?

2 Are translocation and establishment of species technically possible?

3 Do benefits of translocation outweigh the biological and socioeconomic costs and constraints?

Go to options 2 and 3

1 Continue and improve conventional conservation approaches.

2 (i) Improve landscape connectivity in required direction of colonization, (ii) genetically enhance to improve climate robustness of populations within existing geographic range, and (iii) reduce local stressors on population.

3 Invoke ex situ conservation practices (e.g., store egg/sperm/seed).

4 Is it possible to create habitat (e.g., artificial reef, wetlands) at higher latitudes to accommodate "natural" movement?

Will the organisms arrive on their own to new habitat?

5 Wait and facilitate establishment (protect organisms as they arrive).

6 Undertake translocation (assisted migration).

Decision framework for assessing possible species translocation. Assessing the feasibility of whether or not to attempt the movement of a species to prevent its extinction or ecosystem collapse.

PHOTO CREDITS (LEFT TO RIGHT): S. MCINTYRE; JUPITER IMAGES; JAMES P. MCVEY/NOAA SEA GRANT PROGRAM

---

[1] Centre for Marine Studies, Australian Research Council Centre for Excellence in Reef Studies and the Coral Reef Targeted Research Project, www.gefcoral.org, The University of Queensland, St Lucia, Queensland (QLD) 4072, Australia; oveh@uq.edu.au. [2] Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia; lhughes@rna.bio.mq.edu.au. [3] Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) Sustainable Ecosystems, Post Office Box 284, Canberra Australian Capital Territory (ACT) 2601, Australia; Sue.McIntyre@csiro.au. [4] Fenner School of Environment and Society, The Australian National University, Canberra, ACT 0200, Australia; david.lindenmayer@anu.edu.au. [5] Integrative Biology, 1 University Station C0930, University of Texas, Austin, TX 78712, USA; parmesan@uts.cc.utexas.edu. [6] The Ecology Centre, Centre for Applied Environmental Decision Analysis, The University of Queensland, St Lucia, QLD 4072, Australia; h.possingham@uq.edu.au. [7] Department of Biology, University of York, Post Office Box 373, York YO10 5YW, UK; cdt2@york.ac.uk.

*Author for correspondence.

ranges would be dismissed. Some species will also disperse sufficiently to maintain large populations and range sizes (for example, highly dispersive insects or birds with generalist life histories) and others may adapt in situ (*14*). Where species are perceived as being at moderate risk from climate change, improvements in connectivity to actual or potential habitat at higher latitudes and altitudes may be sufficient (*15*).

Moving widespread species within their ranges might, nonetheless, be an important conservation option, especially where significant ecotypic differentiation exists. Moving individuals from "warm-adapted" populations to historically colder locations may increase the probability of subsequent adaptation as the climate changes. For example, staghorn corals (Acroporidae) have wide latitudinal ranges, with low-latitude populations having higher temperature tolerances than those at higher latitudes (*15*, *16*). Populations of staghorn (*Acropora*) corals have already been lost from some high-latitude locations because of increasing thermal stress and declining water quality, and hence, introducing lower-latitude, heat-adapted genotypes to these degraded sites may hold little risk (*16*). Latitudinal and altitudinal clines in genetically based thermal adaptation are equally common on land, e.g., in fruit flies (*17*) and butterflies (*18*). Careful introduction of low-latitude forms of a species may help to preserve it at higher latitude and altitude, as the climate changes.

Assisted colonization should also be considered for species whose ranges have become highly fragmented. Movement in the direction required by climate change may be blocked by human-dominated landscapes [e.g., the endangered Quino checkerspot butterfly, *Euphydryas editha quino* (*19*)]. Dispersal processes that have been disrupted by loss of habitat connectivity could be restored by colonization.

Species that are confined to disappearing habitats present the greatest challenge. Many montane species, for example, face elimination of their habitat as suitable climatic conditions migrate upward and off the top of mountain ranges (*7, 9, 20*). In other cases, the shift of environmental envelopes in a poleward direction may be thwarted by natural barriers (e.g., North African species needing to cross the Mediterranean). In both cases, translocation of species to locations outside their historic range where conditions will be suitable in the medium- to long-term may be the only strategy to prevent extinction.

The assisted colonization of species to a new site depends on additional factors. The first is whether the establishment of species at the target location is technically feasible, and whether the biophysical characteristics of the new location match the needs of the species. In cases where translocation is technically impossible or is prohibitively expensive (*21*), it may be possible to respond by constructing suitable habitat at potential sites for natural colonization. The movement of many coral reef species to higher latitudes, for example, may depend on the presence of benthic structure as opposed to an existing biological community. It might be practical, at small scales, to establish artificial, three-dimensional reef structures ahead of migrating coral, fish, and invertebrate species. On land, it may be possible to restore degraded land with habitats not originally present. Clearly, however, there are financial and other logistic constraints, especially at the scale of the world's ecosystems (*10*, *22*).

One of the most serious risks associated with assisted colonization is the potential for creating new pest problems at the target site. Introduced organisms can also carry diseases and parasites or can alter the genetic structure and breeding systems of local populations. However, most major pest problems have been created by continent-to-continent and continent-to-island translocations or by the transfer of organisms between distinct biogeographic regions within continents (e.g., Nile perch to Lake Victoria). Clearly, risks escalate as species are moved across biogeographical boundaries. Introduction of the cane toad, *Bufo marinus*, from its native range in tropical America to Australia and various tropical parts of the world has been disastrous. This is not the scale of translocation that is being proposed here; we are not recommending placing rhino herds in Arizona or polar bears in Antarctica. We are, however, advocating serious consideration of moving populations from areas where species are seriously threatened by climate change to other parts of the same broad biogeographic region (i.e., broad geographic regions that share similar groups of organisms).

In addition to the ecological risks, socioeconomic concerns must be considered in decisions to move threatened species. Financial or human safety constraints, for example, may make a species' introduction undesirable. It is likely to be unacceptable to move threatened large carnivores or toxic plants into regions that are important for grazing livestock. Ex situ conservation (storage of frozen gametes) may be the only practical option for these species until more suitable habitat can be found or developed in the future.

The reality of a rapidly changing climate has caught many natural-resource managers and policy-makers unprepared. In the past, the assisted migration of a species outside its current range was rarely considered to be an acceptable conservation measure, with the exception of moving species to small, predator- or other threat-free islands (*23*). Larger-scale translocations might now be needed. Consequently, the conservation community needs to move beyond the preservation or restoration of species and ecosystems in situ. Assisted colonization will always carry some risk, but these risks must be weighed against those of extinction and ecosystem loss.

We must contemplate the possibility that some regions of the Earth will experience high levels of warming (>4°C) within the next 100 years, as well as altered precipitation (*10*) and ocean acidity (*8*). Under these circumstances, the future for many species and ecosystems is so bleak that assisted colonization might be their best chance. These strategies will, however, require careful thought and will need to be backed up by detailed scientific understanding if they are to succeed. They must also be accompanied by strategies that address the myriad of other threats in addition to climate change that also endanger species and ecosystems.

References

1. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637 (2006).
2. C. Parmesan, G. Yohe, *Nature* **421**, 37 (2003).
3. L. Hughes, *Trends Ecol. Evol.* **15**, 56 (2000).
4. G. -R. Walther *et al.*, *Nature* **416**, 389 (2002).
5. M. S. Warren *et al.*, *Nature* **414**, 65 (2001).
6. R. Menéndez *et al.*, *Proc. R. Soc. London Ser. B*. **273**, 1465 (2006).
7. D.W. Hilbert, B. Ostendorf, M. S. Hopkins, *Austral. Ecol.* **26**, 590 (2001).
8. O. Hoegh-Guldberg *et al.*, *Science* **318**, 1737 (2007).
9. J. A. Pounds, M. P. L. Fogden, J. H. Campbell, *Nature* **398**, 611 (1999).
10. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the IPCC*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, New York, 2007).
11. M. L. Hunter, *Conserv. Biol.* **21**, 1356 (2007).
12. J. S. McLachlan, J. J. Hellmann, M. W. Schwartz, *Conserv. Biol.* **21**, 297 (2007).
13. J. Rosenhead, in *Rational Analysis for a Problematic World*, J. Rosenhead, Ed. (Wiley, New York, 1989), pp. 193–218.
14. D. K. Skelly *et al.*, *Conserv. Biol.* **21**, 1353 (2007).
15. A. D. Manning, in *Managing and Designing Landscapes for Conservation*, D. B. Lindenmayer and R. J. Hobbs, Eds. (Blackwell Publishing, Oxford, 2007), pp. 349–364.
16. R. Berkelmans, M. J. H. van Oppen, *Proc. R. Soc. London Ser. B*. **273**, 2305 (2006).
17. J. Balanyá *et al.*, *Science* **313**, 1773 (2006).
18. J. G. Kingsolver, K. R. Massie, G. J. Ragland, M. H. Smith, *J. Evol. Biol.* **20**, 892 (2007).
19. C. Parmesan, *Nature* **382**, 765 (1996).
20. S. E. Williams, E. E. Bolitho, S. Fox, *Proc. R. Soc. London Ser. B*. **270**, 1887 (2003).
21. J. Fischer, D. B. Lindenmayer, *Biol. Conserv.* **96**, 1 (2000).
22. M. Scholze, W. Knorr, N. W. Arnell, I. C. Prentice, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 13116 (2006).
23. D. T. Blumstein, *J. Biogeogr.* **29**, 685 (2002).

10.1126/science.1157897

NEUROSCIENCE

Vertebrate Vocalizations

Did the mechanisms underlying vertebrate vocal abilities evolve from a single common ancestor?

Daniel Margoliash and Melina E. Hale

We communicate through speech and language but also with our laughs and cries. Speech and language are learned behaviors that are intimately tied to human cognition. But the more basic pattern is that the structure of vocalization is innately specified; as any parent knows, babies require no practice to cry at birth. Species-specific "calls" are observed in diverse vertebrates and in numerous social contexts, but with a spotty distribution. Many amphibians, as well as some reptiles and fishes, are highly vocal animals. Most birds and mammals call. Did the neural structures for calling among this very broad group of animals arise from a single common ancestor? On page 417 in this issue, Bass *et al*. (*1*) address this question by examining the organization of neural networks underlying vocalization in fish.

One voice. The oyster toadfish (*Opsanus tau*) has a hindbrain and spinal cord organization for vocalization that is similar to that in frogs, birds, and mammals, suggesting a common ancestral neural structure for auditory communication.

CREDIT: HERB SEGARS/ANIMALS ANIMALS

Vertebrates make sounds with an astounding variety of mechanisms, but use only some of these to produce most communication sounds. In terrestrial vertebrates, vocalizations are commonly produced by specialized organs that vibrate air as it moves from the lungs through respiratory tubes. Fishes do not have the vibration-producing structures of terrestrial vertebrates—the larynx in amphibians, reptiles, and mammals or the syrinx in birds—nor do they have an air-filled tube leading to the mouth. Instead, some fishes vocalize with the swim bladder, an air sac used mainly for buoyancy control and as an accessory respiratory organ. The swim bladder is vibrated by associated muscles and acts as a resonance chamber to amplify the sound.

The social contexts of vocalizations are as diverse as their mechanics. Calls are used to communicate across as well as within species, and can be directed at individuals or groups. They are produced in numerous contexts such as aggression, fear, mating, and parent-offspring interactions, signaling intent or surprise (*2*).

Given this substantial diversity in the mechanisms of vocal production and function, one might be tempted to assume that vocal systems arose multiple times across the vertebrates during evolution. An alternative hypothesis, given the existence of vocalizing animals in both of the major clades of Osteichthyes (bony fishes)—the Actinopterygii (ray-finned fishes) and Sarcopterygii (lobe-finned fishes and terrestrial vertebrates)—is that a common origin for vertebrate vocalizations evolved before the divergence of these groups, more than 400 million years ago (*3*).

To determine the relationships among vertebrate vocal systems, Bass *et al*. examined the spatial organization of neurons in the hindbrain vocalization neural network of three species of batracoid fish—the Gulf toadfish (*Opsanus beta*), the oyster toadfish (*Opsanus tau*) (see the figure), and the midshipman fish (*Porichthys notatus*)—and compared them to those of other vertebrates. These fishes are notable for their dramatic calling, used for attracting mates and for territorial defense. The swim bladder muscle used for these vocalizations is one the fastest known vertebrate muscles, contracting at up to 200 Hz (*4*). Such specializations make these species compelling models to study vocalizations in actinopterygian fishes.

Bass *et al*. found that the vocal motor neurons of the toadfish are localized to a region that spans the rostral spinal cord and rhombomere 8, which is the caudal-most segment of the hindbrain. The authors also describe the position of several upstream neuronal elements of the vocalization circuit, including the vocal pacemaker neurons that control rhythmicity. These cells are also located in rhombomere 8, with the pacemaker neurons extending into rostral spinal cord, lateral to the vocal motor neurons. The positions of the vocal motor neurons and other neurons in this circuit are consistent with the organization of the vocal system in frogs, birds, and mammals, supporting the hypothesis that the neural basis of vocal circuits arose before the divergence of sarcopterygian and actinopterygian branches of the vertebrate phylogeny. Encompassing fishes, amphibians, reptiles, birds, and mammals, this is another remarkable example of a conserved vertebrate body plan (*5*).

A single origin for aspects of the neural control of vocalizations leads to other questions about the system, including how much deeper into the history of vertebrates this circuit goes, and how it originated. Addressing such questions is limited by the small number of extant vertebrate outgroups to the bony fishes including Chondrichthyes (sharks, skates, rays, and chimeras) and Agnathans (hagfishes and lampreys), neither of which have air-filled spaces analogous to swim bladders or lungs, nor are known to produce sounds in ways similar to those of lung and swim bladder-based mechanisms. An additional complication is that there is no consistent relation between the vocal motor neurons and the cranial nerve through which they project to muscle. Batracoids and other actinopterygian fishes' motoneurons project through the occipital nerve, whereas birds use the hypoglossal nerve and amphibians and mammals, the vagal nerve (*1*). With current data, the primitive condition of the vocal nerve for Osteichthyes remains unresolved. A complementary approach to examining gross morphology may be to look for molecular markers to help identify homologous circuit elements across diverse taxa.

If vertebrates share a common hindbrain and spinal cord organization for vocalizations, does this also shed light on the evolution of vocal learning? The distribution of learned vocalizations across vertebrates indicates that this behavior has evolved multiple times. Among primates, only humans show compelling evidence of vocal learning. Among mammals, there is increasing evidence for vocal learning, but to date, only in two additional lineages, bats

---

Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA. E-mail: dan@bigbird.uchicago.edu

and cetaceans. Vocal learning is common in birds but has arisen in three separate lineages (songbirds, parrots, and certain hummingbirds), and is not known for more basal birds.

Do these independently evolved behaviors share common neural pathways? At least in songbirds and humans, some of the forebrain pathways implicated in vocal learning share homologous components (*6*, *7*). Also, calling is controlled by a midbrain system that may be common to that of vertebrates (*8*, *9*). It may be that all forebrain mechanisms controlling vocalizations have to interact with the same midbrain system. Such commonalities increase confidence that animal studies can give insight into human vocal learning (*10*, *11*). The story of the evolution of vocalizations is still being written, both for its deep ancestral roots and for its most modern developments.

References

1. A. H. Bass, E. H. Gilland, Robert Baker, *Science* **321**, 417 (2008).
2. A. M. Simmons, A. N. Popper, R. R. Fay, Eds., *Acoustic Communication* (Springer-Verlag, New York, 2003), vol. 16.
3. M. Zhu, *Nature* **410**, 81 (2001).
4. L. C. Rome, D. A. Syme, S. Hollingworth, S. L. Lindstedt, S. M. Baylor, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 8095 (2006).
5. N. Shubin, *Your Inner Fish* (Pantheon Books, New York, 2008).
6. A. Reiner *et al.*, *J. Comp. Neurol.* **473**, 377 (2004).
7. M. A. Farries, D. J. Perkel, *J. Neurosci.* **22**, 3776 (2002).
8. T. J. Seller, *Trends Neurosci.* **4**, 301 (1981).
9. J. M. Kittelberger, B. R. Land, A. H. Bass, *J. Neurophysiol.* **96**, 71 (2006).
10. A. J. Doupe, P. K. Kuhl, *Annu. Rev. Neurosci.* **22**, 567 (1999).
11. Zeigler, P. Marler, Eds., *Behavioral Neurobiology of Bird Song* (New York Academy of Sciences, New York, 2004), vol. 1016.

10.1126/science.1161775

CLIMATE

Was the Younger Dryas Global?

Determining the geographical extent of a 1300-year cold event that occurred just before the current warm period requires accurate chronologies.

Thomas V. Lowell[1] and Meredith A. Kelly[2]

Advance and retreat. According to Ackert *et al.*, eastern outlet glaciers of the Southern Patagonian Icefield, such as the Perito Moreno Glacier shown here, retreated after the Younger Dryas had ended.

CREDIT: MOMATIUK–EASTCOTT/CORBIS

The Younger Dryas cold event, which lasted from 12,900 to 11,600 years ago, was a rapid return to near-glacial conditions during the transition from the Last Glacial Maximum to the current Holocene warm period. Similar rapid climate reversals have been identified during earlier glacial-to-interglacial transitions (*1*). Such millennial-scale climate events may thus play an important role in the transition from glacial to interglacial periods.

The mechanisms that caused the Younger Dryas are not yet fully understood. An important step in understanding these mechanisms is to determine the geographic extent of Younger Dryas–related climate conditions. On page 392 of this issue, Ackert *et al.* (*2*) suggest that Younger Dryas cooling did not influence eastern outlet glaciers of the Southern Patagonian Icefield. Instead, they argue that these glaciers advanced in the early Holocene, likely as a result of changes in southern westerly wind circulation. The absence of Younger Dryas–related cooling in this southern mid-latitude location would have implications for the driving mechanisms of the event.

Uncertainties in dating techniques continue to make it difficult to determine whether mid-latitude locations in the Southern Hemisphere experienced Younger Dryas–related climate changes. To illustrate these uncertainties, consider prior studies and that by Ackert *et al.*, which examine past glacial extents in southern mid-latitude locations. Past glacial extents provide a sensitive proxy for past temperatures (*3*). In general, these extents are dated using either radiocarbon or surface-exposure dating. The latter method is based on the measurement of cosmogenic nuclides in the upper surfaces of boulders on moraines.

However, in New Zealand and South America, where both radiocarbon and surface-exposure methods have been applied to date certain moraines, a distinct offset exists between the resulting ages. Radiocarbon ages in these southern mid-latitude locations indicate that glacial advances occurred before the Younger Dryas. For example, in New Zealand, numerous radiocarbon ages of vegetation overrun by the Franz Josef Glacier show that the Waiho Loop moraine formed prior to, or at the beginning of, the Younger Dryas (*4*). A later study at the same site reports radiocarbon ages that indicate a pre–Younger Dryas glacial advance (*5*). Similarly, in Argentina, one preliminary radiocarbon age indicates that the Perito Moreno Glacier, an eastern outlet of the Southern Patagonian Icefield (see the figure), advanced prior to the Younger Dryas (*6*). In contrast, recently reported surface-exposure (^{10}Be and ^{36}Cl) ages of boulders atop moraines in New Zealand and South America have been interpreted to indicate glacial recession well after the Younger Dryas (*2*, *7*). Both studies conclude that local glacial advances in the southern mid-latitudes were influenced by changes in southern westerly wind circulation in the early Holocene and not by Younger Dryas–related cooling.

There are at least two possible explanations for the offset between radiocarbon and surface-exposure ages of past glacial extents.

---

[1]Department of Geology, University of Cincinnati, Cincinnati, OH 45221, USA. [2]Geochemistry Division, Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA. E-mail: thomas.lowell@uc.edu

First, ages from the two dating techniques may yield accurate ages of different stratigraphic levels. That is, the radiocarbon ages discussed above yield dates of glacial advances, whereas surface-exposure ages yield the time of glacial recession and moraine stabilization. Therefore, an interpretation of the reported ages is that local glaciers in New Zealand and South America advanced prior to the Younger Dryas, remained in extended positions, and receded in the early Holocene. This hypothesis could be tested using temperature and precipitation data from other paleoclimate records as input for glacier modeling studies.

Second, the offset may result from uncertainties associated with the dating techniques. For example, because of plateaus in the radiocarbon calibration curve, reported radiocarbon ages may be uncertain by a few hundred years. Moreover, the uncertainties in cosmogenic nuclide production rates are substantial (as much as 10% at the locations discussed here). One means to address these uncertainties is to use independent dating methods to calibrate site-specific cosmogenic nuclide production rates.

Once again, understanding Earth history hinges on accurate chronologies. As long as these dating issues are unresolved, determining the presence or absence of a Younger Dryas–related climate change at mid-latitudes in the Southern Hemisphere will remain a challenge.

References

1. H. Cheng *et al.*, *Geology* **34**, 217 (2006).
2. R. P. Ackert Jr. *et al.*, *Science* **321**, 392 (2008).
3. J. Oerlemans, *Science* **308**, 675 (2005); published online 3 March 2005 (10.1126/science.1107046).
4. G. H. Denton, C. H. Hendy, *Science* **264**, 1434 (1994).
5. C. S. M. Turney *et al.*, *Quat. Sci. Rev.* **26**, 3037 (2007).
6. J. A. Strelin, E. C. Malagnino, *Quat. Res.* **54**, 339 (2000).
7. T. T. Barrows, S. J. Lehman, L. K. Fifield, P. De Deckker, *Science* **318**, 86 (2007).

10.1126/science.1160148

MATERIALS SCIENCE

Tracking Corrosion Cracking

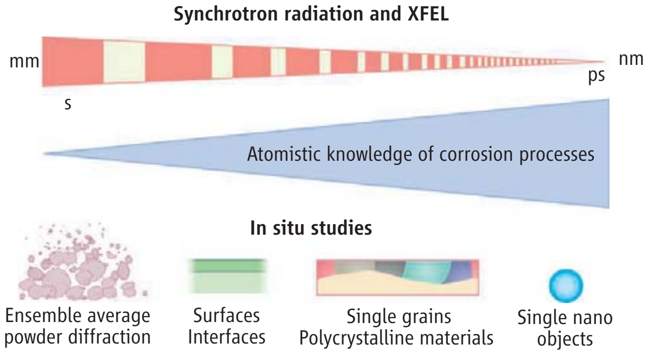
The effects of both mechanical strain and oxidizing agents on grain boundaries in stainless steel can now be visualized.

Andreas Stierle

The corrosion of steel-based mechanical components is an unwanted everyday phenomenon, which destroys about 3% of the annual world gross domestic product (*1*). At the beginning of the 20th century, it was discovered that steels that contained more than 10% chromium are highly resistant to corrosion. Since then, the story of stainless steel has been a surpassing success as a material combining chemical inertness and high mechanical strength for use in applications ranging from kitchen sinks to steam generators. During fabrication and operation, many stainless steel components are exposed to mechanical loads that create high strains inside the material, which results in mechanical failures at unexpectedly low loads in the presence of corrosive agents. As discussed on page 382 of this issue, King *et al.* (*2*) shed light on the microscopic origin of environmental corrosion by studying crack formation in situ, characterizing the effects of stress and corrosive media, such as an acidified solution of $K_2S_4O_6$, on the polycrystalline grain structure of samples in an electrochemical cell.

Grain boundaries are the key players for the combined mechanical and corrosion properties of a material: They serve as sinks for the segregation of impurities inherently included in or added during the manufacturing process. The challenge of studying real polycrystalline samples, as opposed to simpler model samples that are single crystals, is enormous. The internal structure of a polycrystalline material is a nightmare for scientists because it is made up by a three-dimensional (3D) arrangement of individual, more or less randomly oriented crystalline grains of variable size, which interact through their grain boundaries. Conventional (text book) x-ray powder diffraction experiments average over such an ensemble of orientations, which leads to a characteristic, anisotropic broadening of the sharp Bragg diffraction peaks. Retrieving the grain-grain interaction from such Bragg profiles remains a challenge in materials science (*3*). Nevertheless, it is a prerequisite for a full understanding of the mechanical properties of the material at hand.

Only recently has it been shown that highly brilliant, high-energy x-ray synchrotron radiation can be used to analyze the local grain structure of a polycrystalline material, including the local strain state and individual grain orientation under variable external mechanical stress (*4*–*6*). Such a 3D grain structure reconstruction is reported for austenitic stainless steel by King *et al.* (*2*). Synchrotron radiation tomography also allowed King *et al.* to combine grain structure reconstruction with an in situ localization of corrosion processes inside the sample. In these studies, the spatial resolution is limited by the detection system to about 1 μm. The future development of synchrotron beams that are focused to widths of 50 nm (*7*) with stroboscopic methods is expected to deliver even higher-resolution 3D information and a time resolution on the order of 100 fs (*8*) versus 50 to 100 ms at present (see the figure).



A road map for understanding corrosion. Studies performed on technologically relevant materials under realistic conditions in various media (O_2, SO_2, N_2, plasmas, acids, and water) will create a database for improving corrosion resistance. The use of highly focused synchrotron and x-ray free-electron laser (XFEL) radiation should improve both the spatial and temporal resolution of structural studies. The kinetics and dynamics of the corrosion processes can then be revealed at the level of nanocrystalline grains and their atomic interfaces.

CREDIT: ADAPTED FROM ANDREAS STIERLE

During steel welding, a thermally activated carbon segregation process can occur. King *et al.* have found that the grain boundaries in stainless steel, which are structurally less perfect (or where crystallographic planes with high Miller indices meet), are more sensitive to carbon segregation and the formation of chromium carbides, which makes them more sensitive to corrosion.

Max-Planck-Institut für Metallforschung, Stuttgart D-70569, Germany. E-mail: stierle@mf.mpg.de

They showed that this type of grain boundary is responsible for the 3D crack formation and the ultimate mechanical failure of the material. Thus, they have generated insights into the corrosion process at work.

Stress corrosion is a complex phenomenon, and its detailed understanding requires answers to fundamental questions. As a first step, the structure of individual grain boundaries under strain and their contribution to chromium carbide formation must be understood at the atomic level. Second, an atomistic picture of the corrosion process under mechanical load is needed, which may be obtained from future studies on well-defined, atomically controlled interfaces under realistic conditions (*9*). A combination of the x-ray imaging results with transmission electron and atomic force microscopy results (*10*, *11*) and theoretical investigations is highly desirable. The development of better types of stainless steel will require simultaneous improvement of the mechanical properties and corrosion resistance via grain boundary engineering, which will benefit from novel, synchrotron radiation–based analytical tools.

References

1. G. H. Koch *et al.*, Corrosion Cost and Preventive Strategies in the United States. Report FHWA-RD-01-156 (Report by CC Technologies Laboratories, Inc. to Federal Highway Administration (FHWA), Office of Infrastructure Research and Development, McLean, VA, 2001); www.corrosioncost.com/pdf/main.pdf.
2. A. King *et al.*, *Science* **321**, 382 (2008).
3. U. Welzel, E. J. Mittemeijer, *Z. Kristallogr*. **222**, 160 (2007).
4. B. Jakobsen *et al.*, *Science* **312**, 889 (2006).
5. B. C. Larson *et al.*, *Nature* **415**, 887 (2002).
6. S. R. Stock, *Int. Mater. Rev.* **53**, 129 (2008).
7. C. G. Schroer *et al.*, *Appl. Phys. Lett.* **82**, 1485 (2003).
8. C. Blome, Th. Tschentscher, J. Davaasambuu, P. Durand, S. Techert, *J. Synchrotron Radiat.* **879**, 1254 (2007).
9. F. U. Renner *et al.*, *Nature* **439**, 707 (2006).
10. S. M. Bruemmer, L. E. Thomas, *Surf. Interface Anal.* **31**, 571 (2001)
11. H. Yanliang *et al.*, *Mater. Lett.* **62**, 1863 (2008).

10.1126/science.1160939

DEVELOPMENTAL BIOLOGY

Grasping Limb Patterning

Recent studies offer insights into how a molecular signal controls the growth and patterning of the digits on vertebrate limbs during development.

Clifford J. Tabin[1] and Andrew P. McMahon[2]

The secreted molecule Sonic hedgehog (Shh) regulates both the number and type of digits that form on our hands and feet (*1*). Three recent studies appear to challenge the prevailing models of limb patterning (*2*–*4*). However, when viewed in a common context, they fit into a unified view of this process (see the figure).

Although the digits of different vertebrate limbs are dissimilar in the number, shape, and size of bony elements from one species to another, within any species there is little variability in the anterior-posterior (thumb to little finger) pattern. Forty years ago, John Saunders and colleagues showed that information within the posterior region of an embryonic limb bud [the zone of polarizing activity (ZPA)] induces symmetric, mirror-image duplications of the distal limb when grafted to the anterior margin of a second limb bud (*5*). Subsequent studies led to a model in which a graded concentration of ZPA-derived signals coordinates growth (digit number) and pattern (digit identity), the latter by a concentration-dependent mechanism (progressively higher amounts of signals for each more posterior digit).

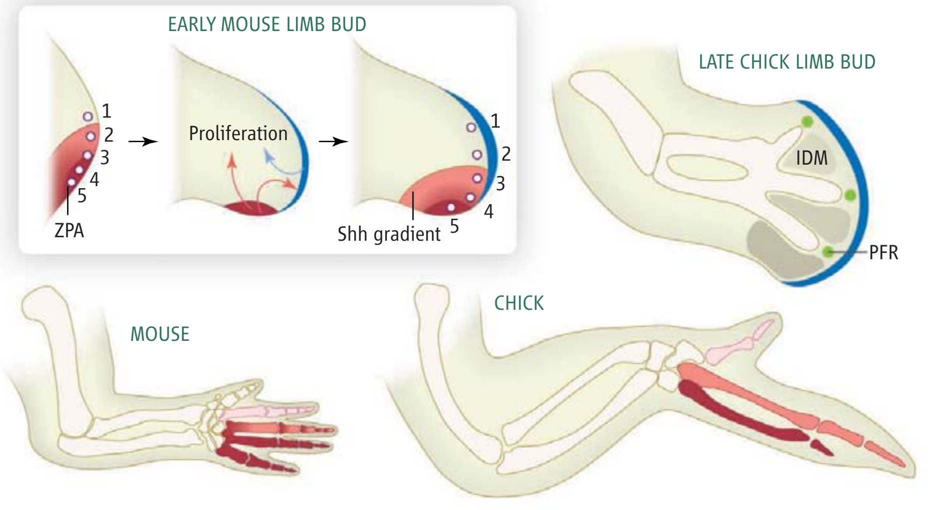
In 1993, the discovery that Shh is the ZPA-derived activity (*1*) enabled mechanism testing. As predicted, Shh forms a gradient across the limb and experimentally is capable of acting as a morphogen, stimulating the formation of distinct digit types in response to different Shh concentrations. But cells also appeared to integrate time of exposure into the equation. A high concentration for a short period led to an anterior digit, whereas continued exposure led to equivalent cells forming posterior digit types (*6*). This also implied that naïve mesenchyme cells are initially instructed as anterior and subsequently "promoted" to posterior digit identities. Further, genetic fate-mapping studies in the mouse, which like a human has five digits, suggested how time of exposure to specific concentrations could be tied to expansion of the mesenchyme (*7*). Skeletal descendants of the Shh-producing ZPA actually comprise a part of the middle digit 3, and much of digits 4 and 5. In the early limb bud, progenitors of all three of these digits express Shh, while at later time points, only the most posterior digit primordium maintains this expression. Thus, cells combine information about the time they reside within the ZPA at maximal Shh concentrations with information on diffusing Shh concentrations perceived after they have exited the ZPA. In the mouse, digit 1 formation is Shh-independent; cells of digit 2 and much of digit 3 rely on diffusing Shh and are specified by low Shh activity; cells of digits 4 and 5 see maximal amounts of Shh and are distinguished from each other because the more posterior digit primoridium has a longer length of ZPA residency.

Recent work from Towers *et al.* (*2*) suggests a model that, at first glance, seems at odds with the above. In the chick wing, only the most posterior digit derives from cells expressing Shh. However, as the chick wing lacks both digits 1 and 5, there is no conflict with previous models regarding the origin and specification of the posterior-most digits. More important, Towers *et al.* indicate that Shh itself is the driver of ZPA expansion, and in a broader sense, acts to integrate growth with patterning. Inhibiting Shh signaling in the early chick limb bud reduces expansion of the limb mesenchyme; fewer digits form, and these are exclusively anterior because the time of exposure is only sufficient for anterior specification. By contrast, when digit progenitor expansion is blocked but signaling retained, the few digits that form are posterior. Consequently, Shh's proliferative function governs exposure time to Shh and progressive specification of digit identity.

The role of Shh as a linchpin linking growth and patterning raises questions about how the morphogen controls cell proliferation. In a number of systems, hedgehog signals directly affect regulators of the cell division cycle, notably N-myc and cyclins D and E (*8*). Towers *et al.* show that expression of these regulatory factors is indeed induced by Shh in the limb, although neither individually is essential for limb development (*9*–*12*). However, in the limb bud, Shh also controls the production of fibroblast growth factors (FGFs) in the apical ectodermal ridge, which overlies the limb mesenchyme (*13, 14*). FGFs can also promote mesenchymal proliferation, raising the possibility of indirect regulatory action through the FGF signaling pathway. Further, Shh does not promote growth following removal of the apical ectodermal ridge, whereas FGFs stimulate proliferation

---

[1] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2] Department of Molecular and Cell Biology, Harvard University, Cambridge, MA 02138, USA. E-mail: tabin@receptor. med.harvard.edu



Digit patterning. Patterning in the early vertebrate limb bud involves integrating total Shh exposure during a dynamic expansion phase. (**Inset**) (Left) Progenitors (numbered 1–5) are exposed to varying amounts of Shh, depending on their position in the ZPA (the source of Shh; maroon), in regions of Shh diffusion (graded red-pink), or in areas beyond the influence of Shh (gray). (Middle) Shh drives expansion of the limb bud, both directly and through expression of FGF in the apical ectodermal ridge (blue). (Right) Digital progenitors are thus displaced relative to the ZPA such that cells are exposed to Shh's patterning influence for varying times. (**Top right**) Exposure to different total amounts of Shh results in differential patterning of the interdigital mesenchyme (IDM), which influences digit growth through BMP signaling in the phalanx-forming regions (PFR). (**Bottom left**) The mouse limb has five digits: two derived from Shh-expressing tissue (maroon), two from lower amounts of Shh (graded red-pink), and one not exposed to Shh. (**Bottom right**) The chick wing has three digits corresponding to the middle three digits of the mouse.

CREDIT: ADAPTED BY P. HUEY/SCIENCE

and outgrowth of the limb in the absence of Shh (*13*). The most likely scenario may be one in which Shh acts to promote growth synergistically with (and only in the presence of) FGF activity. This expands the distal limb and allows space for digits to form, and simultaneously increases the distance between the ZPA and the more anterior digit progenitors as they are progressively patterned by the morphogen.

Surprisingly, another recent study (*3*) by Zhu *et al.* comes to the rather different conclusion that pattern is a *fait accompli* prior to expansion of mesenchymal progenitors. Zhu *et al.* removed Shh activity at different stages of mouse limb development. The earlier the removal, the fewer digits formed. However, the order of digit loss did not reflect their position along the anterior-posterior axis of the limb, but rather, the time of appearance of a given digit condensation (from first to last for Shh-dependent digits: 4, 2, 5, 3). Thus, a digit 3 condensation appears last in the sequence, but digit 3 is the first to be lost upon Shh removal. Whereas the order of digit condensation is quite variable between taxa, this principle of "last in–first out," first proposed by Alberch and Gale (*15*), seems to be a general rule. For example, it is also seen when amphibian limb bud expansion is truncated when exposed to the drug cholcicine (which blocks cell division).

Zhu *et al.* argue that if Shh acts over a long period to progressively pattern more posterior digits, as suggested by Towers *et al.* and earlier studies, this should be reflected in the order of digit reduction rather than the order in which digit condensations are lost. The principle of last in–first out only applies to the location of the digits that are lost. Can one infer digit identity from digit position alone? For example, a short burst of Shh leads to a digit forming in the fourth digit position, but does this have the character of a normal digit 4?

"Digit identity" was coined to draw a parallel (as it turns out, incorrectly) between the specification of different digit morphologies by Hox genes in the vertebrate limb and the mechanism by which segments of a fly (*Drosophila melanogaster*) embryo are established (*16*). It was defined as an empirical property based on the number and relative sizes of the phalanges, but explicitly not on the basis of location within the digit array or shape of the base of the digit. To understand why these latter properties were excluded from the definition, consider complete triphalangeal thumb, the clearest case of a change in digit identity in humans. In this condition, hands appear to have no thumbs and instead present five normal-looking fingers. The first digit has three phalanges (instead of two), is no broader at the tip than the neighboring digit, is located distally in the digital array (instead of set back proximally), and is not rotated into an opposable position (as a true thumb). This is a true transformation to a different identity (a homeotic transformation). Yet the triphalangeal thumb is still in the location of the first digit. Hence, location cannot be a criterion for assigning "identity." In addition, the base of the first metacarpal is still slanted (as opposed to the flat base of the adjacent metacarpal), not because it is incompletely transformed, but because it develops to make an articulation with different wrist elements from its neighbor.

Unfortunately, once location and articulation are eliminated, there is little difference in morphology between different mouse digits with the exception of the non–Shh-dependent biphalangeal digit 1. Zhu *et al.* recognized this problem and tried to circumvent the location argument by examining expression of *Tbx2* and *Tbx3*, transcription factors that demarcate the limits of the digit 3– and digit 4–forming regions, respectively. However, neither gene is essential for normal digit patterning (*17*, *18*). Therefore, their expression may reflect digit location rather than digit identity. Until specific molecular markers of digit identity are isolated (if they exist), the mechanisms underlying digit-type specification are best explored in systems with an unambiguous morphological readout.

Earlier work by Dahn and Fallon challenged the view that digits have an intrinsic identity, suggesting that their morphology is an emergent property governed by adjacent interdigital mesenchyme (*19*). Thus, Shh signaling patterns the interdigital mesenchyme, and this in turn acts as a secondary signaling center to instruct formation of distinct digit types through another signaling molecule, bone morphogenetic protein (BMP). In another recent study, Suzuki *et al.* (*4*) examine how the growth of digit rays—the unsegmented anlagen of the digits—is controlled. Cell-labeling studies in chick embryos demonstrate that mesenchyme beneath the apical ectodermal ridge contributes cells to the distal end of the growing digit ray (the phalanx-forming region). Further, this recruitment of cells correlates with different levels of BMP signaling in the rays of distinct digits (in response to their adjacent interdigital mesenchyme). Although it remains unclear how Shh regulates interdigital mesenchyme properties and how different amounts of BMP signaling translate into different digit types, Suzuki *et al.* provide an important link between positional information provided by Shh activity and digit morphogenesis. Whether similar general principles hold for anterior-posterior patterning of other bones in the limb remains to be determined.

Where do things stand? We know the factor that controls limb formation, but not the detailed mechanism. But these three recent studies provide new insights toward the ultimate goal of a unified framework for understanding how information in a single morphogenetic signal leads ultimately to the generation of a complex pattern.

References

1. R. D. Riddle *et al.*, *Cell* **75**, 1401 (1993).
2. M. Towers *et al., Nature* **452**, 882 (2008).
3. J. Zhu *et al., Dev. Cell* **14**, 624 (2008).
4. S. Suzuki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4185 (2008).
5. J. W. Saunders, M. T. Gasseling, in *Epithelial-Mesodermal Interactions*, R. Fleischmajer, R. E. Billingham, Eds. (Williams and Wilkins, Baltimore, 1968), pp. 78–97.
6. Y. Yang *et al.*, *Development* **124**, 4393 (1997).
7. B. D. Harfe *et al.*, *Cell* **118**, 517 (2004).
8. S. Roy, P. W. Ingham, *J. Cell Sci.* **115**, 4393 (2002).
9. B. R. Stanton *et al.*, *Genes Dev.* **6**, 2235 (1992).
10. J. Charron *et al.*, *Genes Dev*. **6**, 2248 (1992).
11. P. Sicinski *et al.*, *Cell* **82**, 621 (1995).
12. V. Fantl *et al.*, *Genes Dev*. **9**, 2364 (1995).
13. E. Laufer *et al.*, *Cell* **79**, 993 (1994).
14. L. Niswander *et al.*, *Nature* **371**, 609 (1994).
15. P. Alberch, E. Gale, *Evolution* **39**, 8 (1985).
16. C. Tabin, *Development* **116**, 289 (1992).
17. Z. Harrelson *et al.*, *Development* **131**, 5041 (2004).
18. T. G. Davenport *et al.*, *Development* **130**, 2263 (2003).
19. R. D. Dahn, J. F. Fallon, *Science* **289**, 438 (2000).

10.1126/science.1162474

APPLIED PHYSICS

Focus on X-ray Diffraction

The marriage of scanning x-ray microscopy with diffractive imaging provides a route to dramatic increases in spatial resolution.

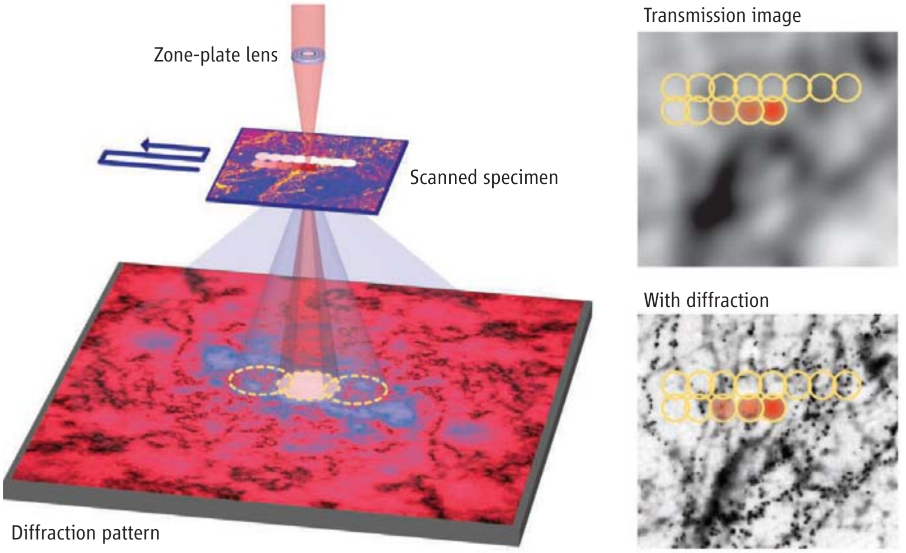
Henry N. Chapman

X-ray microscopy fills the "resolution gap" between light microscopy and electron microscopy. The high penetration of x-rays in matter, however, makes the production of high-resolution lenses challenging. On page 379 of this issue, Thibault *et al.* (*1*) sidestep this problem by recording high-angle (and hence high-resolution) diffraction patterns in the microscope at each point along the specimen. This information is then used to increase the resolution of the image by more than a factor of four. Unlike "lensless" diffractive imaging methods such as crystallography and coherent diffractive imaging (*2*), the scanning diffraction microscopy of Thibault *et al.* offers a general instrument to obtain images, zoom into regions of interest, and build up large fields of view of specimens, within minutes rather than hours.

Scanning diffraction microscopy, or ptychography, was first developed for the scanning transmission electron microscope (STEM) (*3*). The x-ray counterpart, a STXM, is conceptually the same. The x-ray beam is focused onto the sample via a lens, and the transmission is measured in a large-area, single-element detector (see the figure). The image is built up by plotting the transmission as a function of the sample position, as it is rastered across the beam: The smaller the beam focus, the finer the features that can be resolved. If the large-area detector is pixellated, an entire angularly resolved diffraction pattern may be recorded at each point of the scan. Plotting the sum of all detector pixels as a function of beam position simply provides the standard transmission image, but the diffraction pattern contains much more information than that. For example, the pattern can be recorded to high angles, revealing information from features that are much smaller than the beam size. To use this information, we must numerically transport the beams back to the image plane, which is only possible if the phases of the diffracted intensities are also known. This is the "phase problem" of crystallography and other lensless techniques.

Progress has been made in solving the phase problem for a single diffraction pattern recorded from a nonperiodic object (*2*). Such methods, referred to as coherent diffractive imaging (CDI), have achieved better than 10-nm resolution (*4*). The complete complex-valued wave scattered from the object is recovered, giving phase contrast and a way to overcome depth-of-focus limitations (*5*). Although ideal for ultrafast single-shot imaging (*6*), recovering the phases in the diffraction patterns is computationally intensive and requires specific geometric constraints to be satisfied in order to obtain a solution.

For scanning diffraction microscopy, the phase problem becomes tractable: A thin periodic sample will diffract a plane wave into discrete directions (or diffracted orders). For a focused beam consisting of an angular range of plane waves, each diffracted order will now be a disc (dashed white circles in the figure), given by the angular extent (or numerical aperture of the lens). The intensities of the diffracted orders are the same as in the case of illumination with a plane wave, except where they overlap. In these regions of overlap, the



Improved imaging. (Left) Image creation in the scanning transmission x-ray microscope. Simulated images show the improvement of resolution demonstrated by Thibault *et al.* when the angularly resolved transmitted diffraction pattern is used (**lower right**), as compared to the case where total transmission is plotted as a function of beam position (**upper right**).

Centre for Free-Electron Laser Science, University of Hamburg and Deutsches Elektronen-Synchrotron, Notkestrasse 85, 22607 Hamburg, Germany. E-mail: henry.chapman@desy.de

diffracted orders interfere and the intensity depends on the relative phase of the two diffracted orders. It is the folding of one diffraction order onto another that encodes the phase information. As in Gabor holography and “keyhole diffraction” (*7*), there is an ambiguity in the sign of the phase, which scanning diffraction microscopy can resolve from how the diffraction pattern changes as the sample is moved relative to the beam.

The amount of information in the series of diffraction patterns is highly redundant, and many methods to obtain high-resolution images have been proposed and tried over the past 20 years by John Rodenburg and collaborators (*3*). I demonstrated the technique with soft x-rays, recording a total of 4096 diffraction patterns for a single image (*8*), and was delighted in how direct and straightforward it was to obtain phase and amplitude images. However, the long readout time of the charge-coupled device detector between each diffraction pattern exposure discouraged further research.

Two key developments have created the breakthrough demonstrated by Thibault *et al.* The first is the Pilatus detector (*9*), a large-area detector using advances from high-energy physics experiments. Each pixel is activated independently and counts up the charge as the photons arrive, speeding up the detector and eliminating noise. The second is the adaptation of the iterative methods from CDI (*10,11*), which are stable when applied to the redundant ptychographic data sets.

These two improvements should be implemented at every STXM at synchrotrons worldwide. Doing so will be revolutionary, allowing desktop computers to overcome x-ray optical limitations to reach resolutions below 10 nm. This is not to say that work in x-ray optics should cease. On the contrary, focusing the x-ray beam to dimensions less than the unit cell size of a macromolecular crystal may indeed be feasible for hard x-rays (*12*). In this case, with thin crystals it will be possible to overlap the Bragg peaks and read out their phases, thereby achieving atomic-resolution x-ray microscopy.

References

1. P. Thibault *et al.*, *Science* **321**, 379 (2008).
2. J. C. H. Spence, in *The Science of Microscopy 2nd Edition*, P. W. Hawkes J. C. H. Spence, Eds. (Springer, New York, 2007), pp. 1196–1127.
3. J. M. Rodenburg, in *Advances in Physics and Electron Physics*, vol. 150, P. W. Hawkes, Ed. (Elsevier, San Diego, 2008), pp. 87–184.
4. J. Miao *et al.*, *Phys. Rev. B* **67**, 174104 (2003).
5. H. N. Chapman *et al.*, *J. Opt. Soc. Am A* **23**, 1179 (2006).
6. H. N. Chapman *et al*, *Nat. Phys.* **2**, 839 (2006).
7. B. Abbey *et al.*, *Nat. Phys.* **4**, 394 (2008).
8. H. N. Chapman, *Ultramicros.* **66**, 153 (1996).
9. C. Broennimann *et al.*, *J. Synchr. Rad.* **13**, 120 (2006).
10. V. Elser, *J. Opt. Soc. Am. A* **20**, 40 (2003).
11. J. M. Rodenburg, H. M. L. Faulkner, *Appl. Phys. Lett.* **85**, 4795 (2004).
12. H. C. Kang *et al.*, *Phys. Rev. Lett.* **96**, 127401 (2006).

10.1126/science.1161183

CLIMATE CHANGE

A Hotter Greenhouse?

A controversy about how high tropical temperatures were during a previous hot period may soon be resolved.

Matthew Huber

Scientists have long been puzzled by the fact that mid-to-high latitude continental interiors and the poles in the Eocene (55 to 34 Million years ago) were much warmer than today, without freezing winters (*1*), while tropical sea surface temperatures (SSTs) were apparently near-modern (*2*, *3*). Mechanisms proposed by climate modelers to maintain high-latitude warmth require substantial tropical temperature increases (*4*, *5*). The implication is that fundamental gaps remain in our understanding of climate dynamics. Many hypotheses have been advanced to resolve this paradox: Paleoclimate proxies require reinterpretation, boundary conditions need improvement, or a major mechanism is missing from climate models. However, no proposal has led to a simple, general solution. A resolution might be in sight based on efforts to develop better climate proxies and multiproxy Eocene records. But this resolution may present new challenges.

Most of what we know about Eocene SSTs comes from records of the oxygen isotopic composition ($\delta^{18}O$) of planktonic foraminiferal calcite shells that reflect the temperature of shell formation and other factors (*3*, *5*). Reconstructing temperatures using $\delta^{18}O$ of foraminifera requires knowing or assuming the $\delta^{18}O$ of the seawater in which they grew. This composition changes globally as terrestrial ice volume varies, and regionally with changes in net evaporation. Estimated evaporation is, in turn, affected by SST. Interpretation is complicated by ambiguity about the depth range at which foraminifera were calcifying and by seasonal biases.

Even more troubling, the $\delta^{18}O$ of planktonic foraminiferal shells may be diagenetically altered once they reach the seafloor (that is, in the early phases of burial, secondary calcite can precipitate from the much colder sediment pore-water fluids onto or into the shell itself). Alteration can push SST estimates toward cooler (bottom water) values (*6*). Alteration is enhanced by increased rates of pore-water flow and increased burial temperature and pressure (*6*). Thus, the best preservation is expected from relatively impermeable sediments that have not been buried deeply or exposed to high temperatures.

Pearson *et al.* (*7*) sought and found records from clay-rich regions in the hope of identifying whether previous records were biased. In a series of pioneering studies (*7–10*), they recovered foraminifera from shallowly buried impermeable clays in several regions, most notably in Tanzania. When viewed under a light microscope, the recovered foraminifera were translucent or “glassy,” as modern samples are, unlike the “frosty” or “chalky” appearance of typical foraminifera used in most previous analyses. Under a scanning electron microscope, the frosty shells were revealed as recrystallized. Analysis of $\delta^{18}O$ in the glassy foraminifera indicated temperatures warmer by 5° to 10°C than previous reconstructions (*8*). The implications are that the shells in many reconstructions from open-ocean sediments were altered and that much of the original isotopic SST signal had been overprinted by cold deep water trends. Acceptance of this viewpoint is, however, not universal (*11*, *12*). Focus has shifted to using other proxies, such as Mg/Ca and TEX_{86}, to develop independent SST records.

The Mg/Ca ratio of foraminiferal shells is a function of temperature and offers a paleoclimate record independent of seawater $\delta^{18}O$. However, the exact relationship is species-specific; depth, seasonality, and alkalinity have influences, the global value of seawater Mg/Ca must be measured or modeled as it varied in time, and Mg/Ca values may also be diagenetically altered, although perhaps to a lesser degree than $\delta^{18}O$ (*8*).

Earth and Atmospheric Sciences, Purdue University, West Lafayette, IN 47906, USA. E-mail: huberm@purdue.edu

By careful site, sample, and calibration selection, these uncertainties can be reduced, but the global seawater Mg/Ca ratio is relatively unconstrained, imposing temperature uncertainties of ±4°C during the Eocene. For reasonable choices of seawater Mg/Ca composition, the temperature agreement between δ^{18}O and Mg/Ca is highest in glassy foraminifera (*8*), indicating that they are more likely to represent the true SST. Furthermore, Mg/Ca-derived SSTs are much warmer than those derived from the δ^{18}O of frosty or chalky (presumably altered) foraminifera in open ocean sites (*8*, *13*). Unfortunately, comparison at individual sites of δ^{18}O and Mg/Ca still reveals discrepancies of 2° to 9 °C in glassy foraminifera, giving a plausible range of SST from 25° to 34°C at one site (*8*). Clearly, another independent proxy is needed to further improve estimates.

TEX_{86} is a promising new proxy for mean annual SST, based on organic molecules derived from crenarcheota (*14*). The technique is independent of surface freshwater balance and seawater composition. Pearson *et al.* used this approach in their Tanzanian record, and their three proxy records showed impressive congruence (*8*, *9*), providing a more credible interpretation than any single proxy. But TEX_{86} is imperfect (*15*) and novel enough that important kinks have not been worked out (*16*). The method has been difficult to calibrate at the hot (>30°C) temperatures central to Early Eocene climate paradoxes. Various studies, including that of Pearson *et al.* (*9*), used a previously developed ad hoc calibration using different slopes at high and low temperatures, clipping maximum temperatures.

Recent incubation (*14*) and core-top (*15*) studies resulted in a new calibration for TEX_{86} that is linear up to 40°C, which raises interpreted peak tropical SST by ~5°C from those originally published using TEX_{86} (*9*). Thus, a newer interpretation (see supporting online material) for the warmest Eocene suggests tropical SSTs in the 35° to 40°C range, not the 33° to 28°C range published in 2007 (*9*), or the 25° to 30°C range as thought a decade ago (*3*), or the 20° to 25°C range accepted two decades ago (*2*).

Half of Earth's surface area is in the tropics, so changes and uncertainties in tropical temperatures dominate any climate sensitivity estimate. If SSTs were truly ~35°C at times in Tanzania (19°S) or New Jersey (~30°N), some tropical regions must have been much hotter. This has thought-provoking implications for paleoclimate, vegetation, and carbon cycle evolution.

Hot tropics. Recent studies suggest that during the Eocene, tropical temperatures were much higher than they are today. The resulting heat stress may have led to tropical vegetation die-offs, with profound implications for climate and the carbon cycle.

CREDIT: FRANS LANTING/CORBIS

First, tropical temperatures above 31°C offer no evidence for a climate thermostat, that is, a strict mechanism that maintains tropical SSTs in the modern range; climate dynamicists trying for decades to explain thermostats may have been chasing a chimera. Second, climate models might be able to reproduce warm poles and warm extratropical continental winters, given that these new tropical SSTs imply closer to modern temperature gradients (*5*).

Third, during the warmest parts of the past 65 million years—that is, the Paleocene-Eocene Thermal Maximum (PETM) and subsequent brief, sudden "hyperthermal" phases of the Early Eocene Climate Optimum (*17*)—tropical vegetation may have been above the upper limits of its thermal tolerance (*18*). Most plants, especially the C3 plants that comprised Eocene floras, have physiological mechanisms that break down in the 35° to 40°C range (*18, 19*); in particular, they can die because photorespiration dominates over photosynthesis (*18–20*). Annual mean temperatures greater than 35°C can be plausibly reconstructed to have been widespread equatorward of 35° latitude (*8*, *9*, *21*, *22*), so floras may have been thermally stressed, and perhaps undergoing water stress in the warmest intervals. There is some evidence of tropical floral extinctions during the warmest periods (*23*, *24*), while forests thrived at higher latitudes.

This scenario may be a missing link in the hypothesis (*25*) that carbon cycle and climate changes during the PETM were caused by oxidation of the terrestrial biosphere. It is well established that a major tropical vegetation die-off in a global warming world has profound temperature, precipitation, and carbon feedbacks (*20*). Carbon cycle modeling (*26*) suggests that the terrestrial carbon pool could have been much larger than modern, ~6000 gigatons of carbon. The gradual warming preceding the PETM may have loaded a terrestrial carbon storage gun, and crossing the 35°C threshold may have triggered it. Tropical die-back after an initial warming (*22*) might have added thousands of gigatons of carbon into the atmosphere and further increased temperatures by radically reducing evapotranspirative fluxes that normally cool tropical landmasses. Tropical heat death helps resolve two mysteries: the magnitude of the carbon and climate excursion at the PETM, and the fact that these abrupt warmings occur during broader intervals of extreme warmth, rather than in cold intervals as expected from methane degassing (*27*).

The recent results suggest that, rather than being a stable cradle for tropical life, the tropics may have been a crucible; during warming, many taxa may have been forced to flee poleward, innovate, or face extinction (*28*). These far-ranging implications are a lot to place on the narrow shoulders of the few published proxy records, but they highlight the importance of the next challenge: collecting more tropical multiproxy records and establishing the accuracy of existing ones.

References

1. D. R. Greenwood, S. L. Wing, *Geology* **23**, 1044 (1995).
2. E. J. Barron, *Paleoceanogr.* **2**, 729 (1987).
3. T. J. Crowley, J. C. Zachos, in *Warm Climates in Earth History*, B. T. Huber, K. G. MacLeod, S. L. Wing, Eds. (Cambridge Univ. Press, Cambridge, UK, 2000), pp. 50–76.
4. C. Shellito *et al.*, *Palaeogeog. Palaeoclimatol. Palaeoecol.* **193**, 113 (2003).
5. M. Huber, L. C. Sloan, *Paleoceanogr.* **15**, 443 (2000).
6. D. P. Schrag, *Chem. Geol.* **161**, 215 (1999).
7. P. N. Pearson *et al.*, *Nature* **413**, 481 (2001).
8. P. F. Sexton *et al.*, *Geosystems* **7**, Q12P19 (2006).
9. P. N. Pearson *et al.*, *Geology* **35**, 211 (2007).
10. C. H. Lear *et al.*, *Geology*, **36**, 251 (2008).
11. J. C. Zachos *et al.*, *Nature,* **419**, 897, (2002).
12. A. Dutton, T. J. Bralower, *Geology*, e.152 , (2008).
13. A. K. Tripati *et al.*, *Paleoceanogr.* **18**, 1101 (2003).
14. S. Schouten *et al.*, *Org. Geochem*. **38**, 1537 (2007).
15. J.-H. Kim *et al.*, *Geochim. Cosmochim. Acta* **72**, 1154 (2008).
16. C. Turich *et al.*, *Geochem. Cosmochem. Acta* **71**, 3272 (2007).
17. J. C. Zachos *et al.*, *Nature* **451**, 279 (2008).
18. J. Berry, O. Bjorkman, *Ann. Rev. Plant. Physiol.* **31**, 491 (1980).
19. T. D. Sharkey, *Science* **287**, 435 (2000).
20. H. D. Matthews *et al.*, *Glob. Biogeoch. Cyc.* **21**, GB2012 (2007).
21. J. C. Zachos *et al.*, *Geology* **34**, 737 (2006).
22. A. Sluijs *et al.*, *Nature* **450**, 1218 (2007).
23. G. Harrington, C. A. Jaramillo, *J. Geological Soc.* **164**, 323 (2007).
24. C. A. Jaramillo *et al.*, *Science* **311**, 1893 (2006).
25. A. C. Kurtz *et al.*, *Paleoceanography* **18**, 14-1 (2003).
26. J. D. Beerling, *Paleogeogr. Paleoclim. Paleoecol.* **161**, 395 (2000).
27. M. Pagani *et al.*, *Science* **314**, 1556 (2006).
28. J. J. Tewskbury *et al.*, *Science* **320**, 1296 (2008).

Supporting Online Material

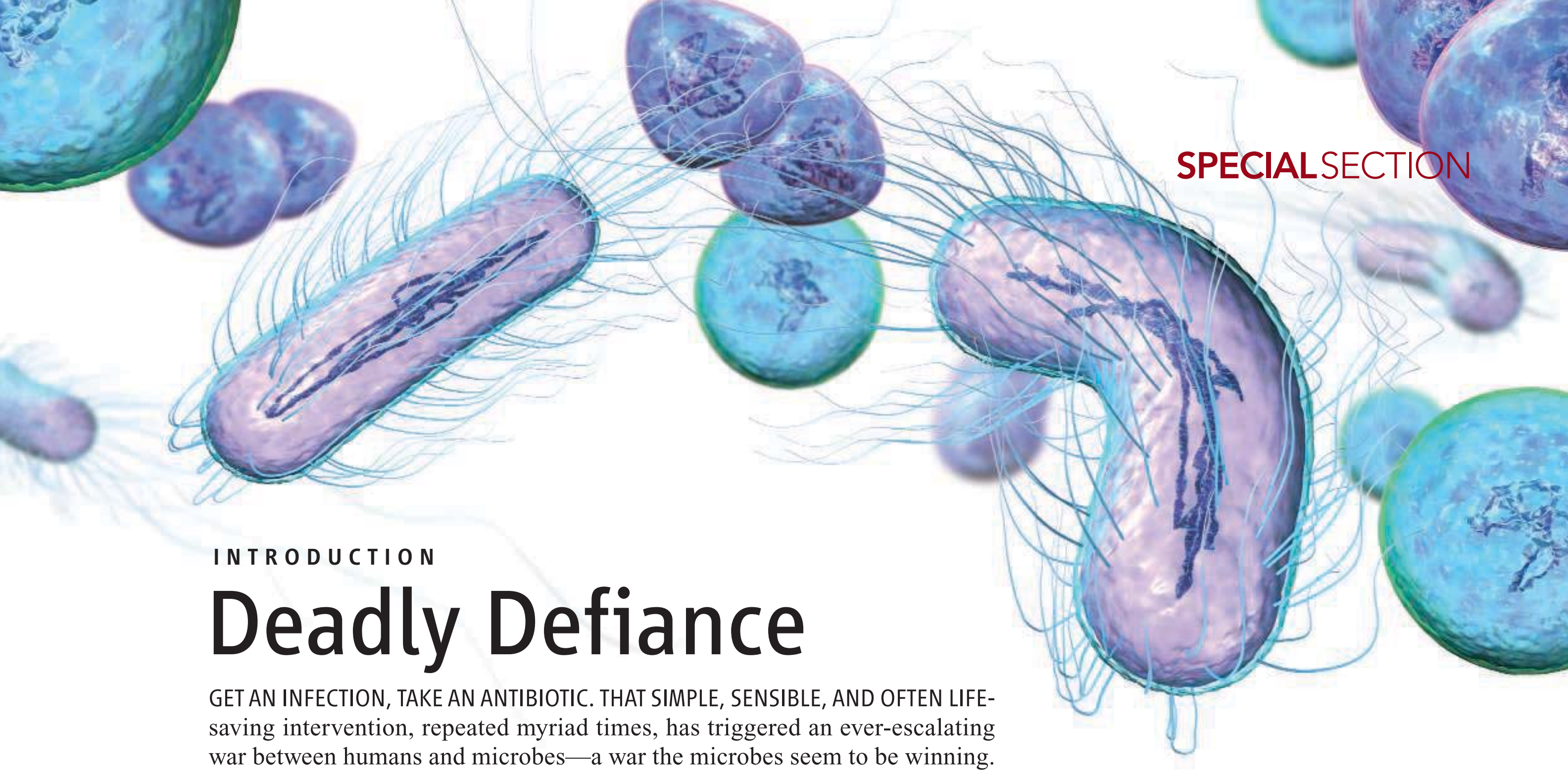
www.sciencemag.org/cgi/content/full/321/5887/353/DC1

Fig. S1

References

10.1126/science.1161170

Published by AAAS



CREDIT: C. BICKEL/SCIENCE

INTRODUCTION

Deadly Defiance

GET AN INFECTION, TAKE AN ANTIBIOTIC. THAT SIMPLE, SENSIBLE, AND OFTEN LIFE-saving intervention, repeated myriad times, has triggered an ever-escalating war between humans and microbes—a war the microbes seem to be winning. Almost as soon as penicillin was introduced in 1942, the bacteria it was designed to defeat began evolving to resist it. Now many common bacteria have acquired resistance to multiple antibiotics, making some infections extraordinarily difficult, if not impossible, to treat.

So far, most of these pernicious drug-resistant infections have been confined to hospitals, where opportunities abound for resistant bacteria to spread and enter the bloodstream or infect open wounds. But the recent emergence of multiple-antibiotic–resistant bacteria in the broader community, particularly methicillin-resistant *Staphylococcus aureus* in the late 1990s, has sparked considerable alarm, not to mention doomsday scenarios of untreatable bacterial infections arising from everyday scrapes and scratches.

In a News story in this special issue, Taubes chronicles the predictable, if unnerving, rise and spread of so-called bad bugs. As the number of antibiotic-resistant bacteria has increased, drug companies have been fleeing the field, leaving a dearth of new antibiotics, especially for the hard-to-kill Gram-negative bacteria. Short of a new wonder drug, the only near-term fix is to curb antibiotic use. Taubes describes several strategies, such as encouraging a switch from broad- to narrow-spectrum antibiotics and investigating the benefits of a shorter course of antibiotics instead of the standard 7 to 10 days.

In a second News story, Marshall goes behind the scenes with Partners in Health (PIH), a nonprofit organization associated with Harvard University's Brigham and Women's Hospital, in Tomsk, Siberia, visiting prisons, hospitals, and other hotbeds of multidrug-resistant tuberculosis (TB). Collaborating with local authorities, PIH has launched an innovative program for tackling this growing threat in places such as Siberia, where resources are limited. This includes roving community workers or nurses who track down recalcitrant patients and give them their medicine, as well as trying to optimize the use of existing (and outdated) drugs. PIH's persistence and vigilance have paid off, but it's unclear whether such a labor-intensive model can work elsewhere. Meanwhile, even in Tomsk, the percentage of TB cases that are drug-resistant remains high—not an auspicious sign.

Two Perspectives explore how resistance emerges and whether such mechanisms reveal possible new interventions. On p. 365, Martínez discusses the evolution of antibiotic resistance genes in bacteria in natural environments and considers whether this might offer clues for fighting infections in more familiar clinical settings. Monk and Goffeau (p. 367) look at modes of fungal drug resistance and whether broad-spectrum fungicides, yet to be developed, might be a solution.

–LESLIE ROBERTS AND STEPHEN SIMPSON

Drug Resistance

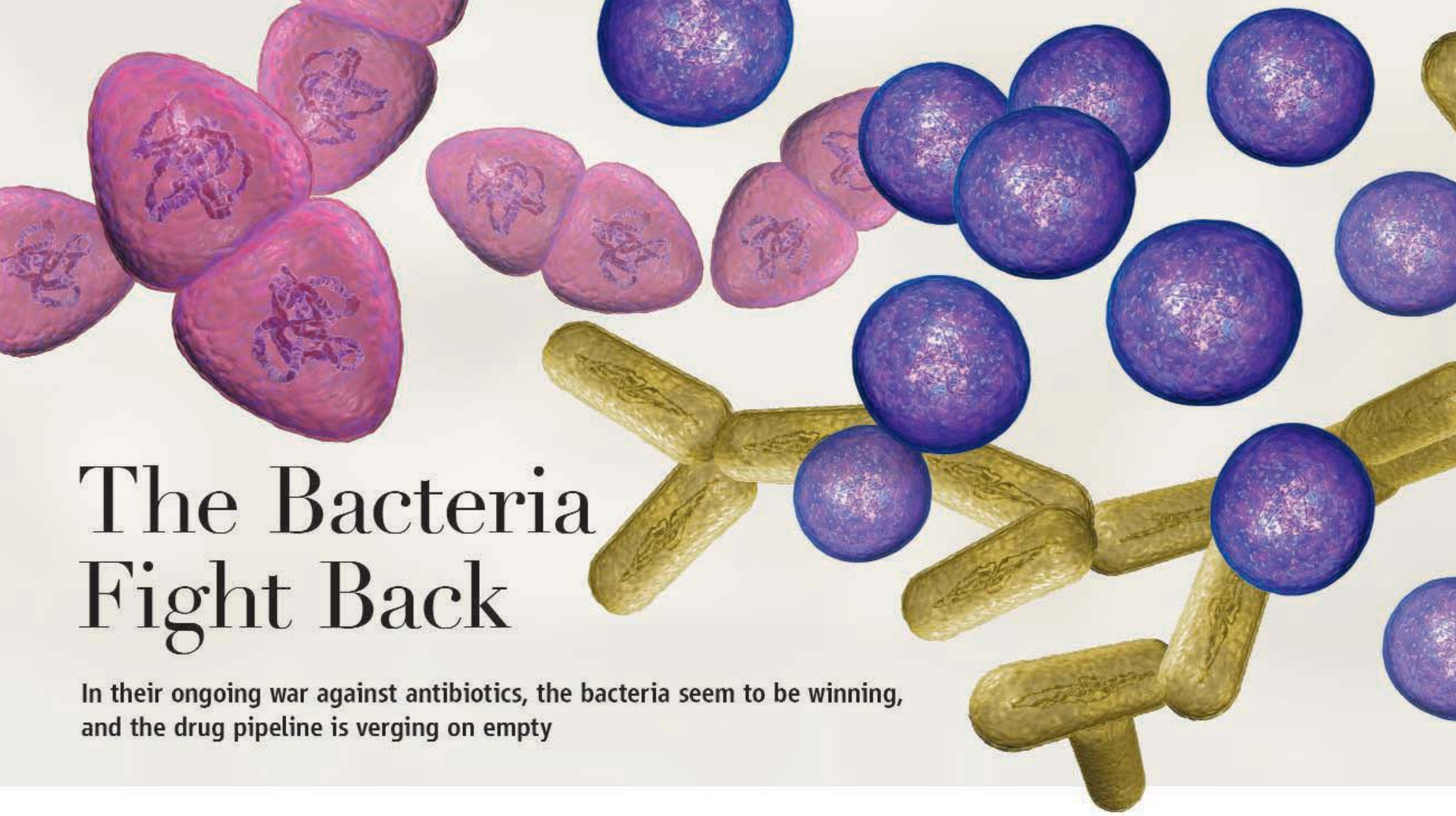
CONTENTS

News

Perspectives

See also related Science *Podcast*

Published by AAAS



The Bacteria Fight Back

In their ongoing war against antibiotics, the bacteria seem to be winning, and the drug pipeline is verging on empty

MAYBE IT WAS JUST A BAD MONTH—AN unfortunate statistical fluctuation. Maybe not. As Vance Fowler, an infectious-disease specialist at Duke University Medical Center in Durham, North Carolina, tells it, the first case appeared in early spring 2008: a 13-year-old girl whose bout with the flu evolved into a life-and-death struggle, still ongoing, with necrotizing pneumonia and a particularly pernicious strain of bacteria known as methicillin-resistant *Staphylococcus aureus* (MRSA). Should the girl survive, her life will be "forever changed," says Fowler, from pulmonary disease caused by the death of the lung tissue. The next case, a week or so later, was a research technician from Fowler's laboratory, admitted to the hospital with a facial abscess that showed no signs of healing. Again, MRSA was the cause. A week or so after that, the victims were a husband and wife. "Both were admitted with life-threatening acute MRSA infections out of nowhere," he says. "Multiple surgeries. Life- and limb-threatening infections." Neither one worked in a hospital or a long-term care facility, the kind of environments in which such bacteria might commonly be found. Nor had they visited one recently. So how did they get it? "Bad luck, bad genes, a bad bug, or all three," says Fowler.

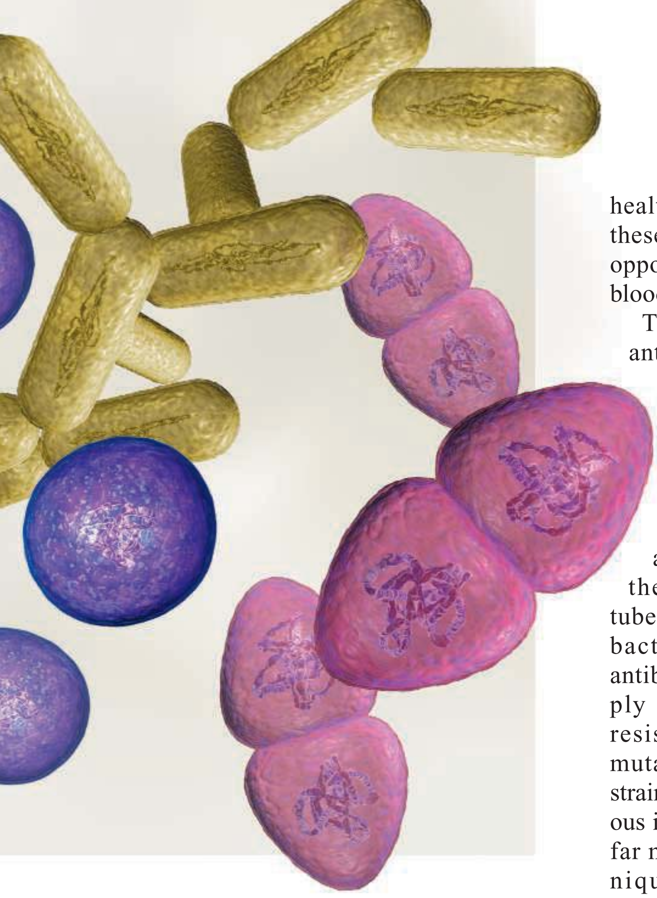
The last decade has seen the inexorable proliferation of a host of antibiotic-resistant bacteria, or bad bugs, not just MRSA but other insidious players as well, including *Acinetobacter baumannii*, *Enterococcus faecium*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Enterobacter* species. The problem was predictable—"resistance happens," as Karen Bush, an anti-infectives researcher at Johnson and Johnson (J&J) in Raritan, New Jersey, puts it—but that doesn't make it any easier to deal with. In 2002, the U.S. Centers for Disease Control and Prevention (CDC) estimated that at least 90,000 deaths a year in the United States could be attributed to bacterial infections, more than half caused by bugs resistant to at least one commonly used antibiotic. Last October, CDC reported in the *Journal of the American Medical Association* that the number of serious infections caused by MRSA alone was close to 100,000 a year, with almost 19,000 related fatalities—a number, an accompanying editorial observed, that is larger than the U.S. death toll attributed to HIV/AIDS in the same year.

Essential but not enough. Washing hands is one step, but ridding a hospital of resistant bacteria also requires identifying and isolating infected patients.

So far these outbreaks have been concentrated in hospitals, where the environment is particularly conducive to the acquisition and spread of drug-resistant bugs. But the big worry, for Fowler and others, is that they will spread to the wider community—a nightmare scenario, he says. MRSA is particularly worrisome, but so is another class of bacteria, called Gram-negative bacteria, that are even tougher to defeat. These include *A. baumannii*, which has plagued injured soldiers returning from Iraq. For these bacteria, the pipeline of new antibiotics is verging on empty. "What do you do when you're faced with an

CREDITS (TOP TO BOTTOM): (ILLUSTRATIONS) C. BICKEL/SCIENCE; LAUNETTE FLORIAN /LANDOV

Published by AAAS

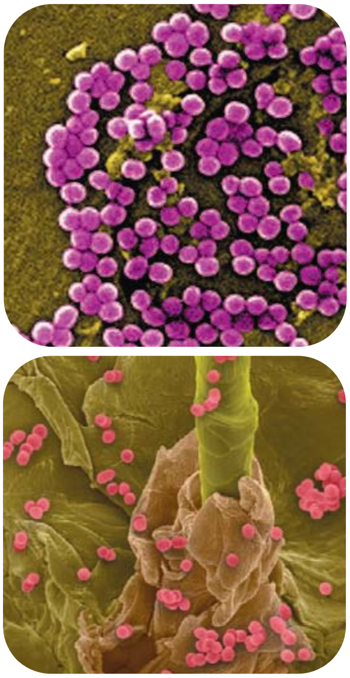


infection, with a very sick patient, and you get a lab report back and every single drug is listed as resistant?" asks Fred Tenover of CDC. "This is a major blooming public health crisis."

Right bug, wrong place

One of the many misconceptions about bacterial infections is that the bugs involved are not native to the human body or are particularly pernicious to begin with. Virtually all bacteria are capable of causing serious infections, at least in immunocompromised patients, although most do not. In hospitalized patients, many infections arise from the patient's own bacterial flora, flourishing where they're not supposed to be. Pneumonia, for instance, can be caused when bacteria from the mouth are aspirated to the lungs. Just as *Escherichia coli* is a normal inhabitant of the gut, *S. aureus* colonizes the skin and mucosal surfaces in the nose in 30% of the population. When it sets up shop somewhere else, *S. aureus* can cause a host of infections, including skin abscesses, necrotizing pneumonia, joint infections, and heart valve infections known as endocarditis. Similarly, *S. epidermidis* normally colonizes the human skin, but when it gets into the bloodstream, it can cause sepsis and endocarditis, as well as infections involving prosthetic devices such as pacemakers and artificial joints. The risk of acquiring one of these serious infections is highest in hospitals and health-care facilities simply because these environments offer the greatest opportunities for bacteria to enter the bloodstream or infect open wounds.

Treating a bacterial infection with antibiotics is the obvious first step. But in the 65 years since the first widespread use of penicillin during World War II, infectious-disease specialists have been treated to an ongoing tutorial in the many ways bacteria can acquire and spread resistance to these drugs. Unlike tuberculosis and other bacteria, in which antibiotic therapy simply selects for rare resistance-bestowing mutants, the bacterial strains that are so insidious in hospitals employ far more diverse techniques, says Louis Rice of Case Western Reserve University and the Louis Stokes Cleveland VA Medical Center in Ohio. *S. aureus* and *Enterococcus*, for instance, can acquire resistance by exchanging entire genes or multiple genes with other bacteria, either through plasmids—loops of DNA that are independent of the bacterial chromosomal DNA—or so-called gene cassettes or transposable elements that can be inserted directly into the chromosomal DNA.



Bad actors. Methicillin-resistant *S. aureus* (*above*) and vancomycin-resistant *Enterococcus*.

Penicillin and all penicillin-like antibiotics are ringlike molecular structures, known technically as β-lactams. They work by attacking a particular cell wall enzyme in the bacteria. The first strains of penicillin-resistant *S. aureus*, which arose within a few years of penicillin's introduction, were strains that have a survival advantage because they naturally produce an enzyme—penicillinase, one of a class of enzymes known as β-lactamases—that destroys the ring structure of penicillin. Within a decade, the effectiveness of penicillin against hospital-acquired staph infections was "virtually annulled," says microbiologist Alexander Tomasz of Rockefeller University in New York City, by "plasmid epidemics" that then spread the penicillinase gene through the entire species of *S. aureus*.

The pharmaceutical industry responded in the 1950s with a host of semisynthetic penicillins designed to be resistant to penicillinases. Methicillin, introduced in 1959, was believed to be the most effective. As Graham Ayliffe, a veteran hospital infection expert at the University of Birmingham in the U.K., recalled, "this was [supposed to be] the end of the resistant staphylococcus." Within 2 years, however, hospitals in Europe were identifying strains of *S. aureus* that were resistant to methicillin: the first MRSA strains.

Researchers later realized that these strains had taken a different route to acquiring resistance. Rather than generating new or different β-lactamases, which could attack the antibiotic directly, they had acquired a new gene entirely, called *mec*A, that coded for a variant of the antibiotic's target: the penicillin-binding protein. When the antibiotics attack the original penicillin-binding protein, explains Tomasz, this "surrogate" binding protein "takes over the task of cell wall synthesis" and works to keep the antibiotic at bay. The *mec*A gene itself, says Tomasz, appears to derive from a common bacterium on the skin of domestic and wild animals known as *S. sciuri*. How *S. aureus* came to acquire the gene is a mystery, but since it did, it passes it on by exchanging entire gene cassettes with the *mec*A gene on them.

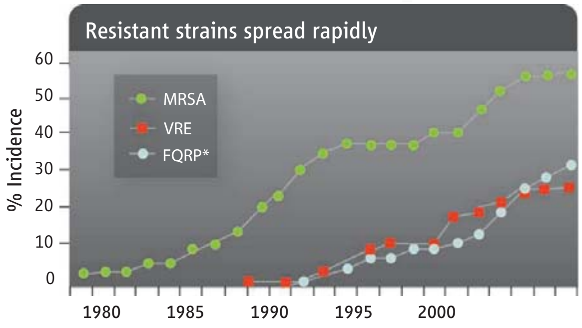
Breaking out

Through the 1970s and 1980s, MRSA remained little more than a nuisance bug, although occasional hospital outbreaks would have to be reined in with strict isolation and control programs. In the mid-1980s, typically only 1% to 5% of all *S. aureus* isolates were methicillin-resistant, says Henry Chambers, an infectious-disease specialist at the University of California, San Francisco. Around that time, *S. aureus* began to acquire genes that confer resistance to other common antibiotics,

CREDITS (TOP TO BOTTOM): JEFF HAGEMAN, M.H.S./CDC; VISUALS UNLIMITED/CORBIS

Published by AAAS

apparently from methicillin-resistant *S. epidermidis* and carried on the same mobile cassettes as *mec*A. The result was a bug that was both far more difficult to treat and, as Chambers says, “pretty adaptive to surviving in hospitals.” Today, 60% to 70% of all *S. aureus* strains found in hospitals are multidrug-resistant MRSA.



*Fluoroquinoline-resistant *Pseudomonas aeruginosa*

Worries intensified when MRSA appeared a decade ago as a community-acquired infection rather than one exclusive to health-care settings. In 1999, CDC reported on four deaths in Minnesota and North Dakota, all children, all caused by MRSA infections that could not be traced back to hospitalizations by either the patients or family members. Somehow the *mec*A gene had emerged in *S. aureus* strains outside hospitals or health-care facilities. “This was a real biological success story,” says CDC’s Tenover. “And it all happened off our radar screens.” MRSA isolates then began to appear in a range of unexpected community settings: children in day-care centers, army recruits, athletes in contact sports, native Americans living on reservations, prison populations, intravenous drug users, and among men who have sex with men.

The possibility that these MRSA strains were simply hospital strains that had migrated out into the community was refuted by analysis of the gene cassettes carrying the resistance. In hospital strains, these cassettes are relatively large and carry multiple resistance-bestowing genes, explains Tomasz. The “oddball” cassettes carrying the *mec*A gene in the early community-acquired isolates were small and contained only the one gene. In the last few years, however, MRSA strains in the community have begun to acquire multidrug resistance, suggesting that they’ve been intermingling with the hospital strains.

A half-dozen community-acquired MRSA clones have now spread around the world as their prevalence in the community has continued to increase; in San Francisco, for instance, up to 50% of all *S. aureus* isolates outside health-care settings are now methicillin-resistant. “These methicillin-resistant strains seem to be replacing the susceptible strains of *S. aureus* in the general population,” says Mark Enright, an infectious-disease specialist at Imperial College London (ICL), “which means people are carrying strains of MRSA in their nose in the community. Now when they get infections, ones that were formerly treatable are going to be replaced with difficult-to-treat infections.”

The public health anxiety increased still further in 2002 with the detection of isolates of MRSA that were fully resistant to the antibiotic vancomycin, traditionally considered the last resort for treating resistant staphylococcal infections. These *S. aureus* isolates seem to have acquired a gene for vancomycin resistance—*van*A—from enterococci, and specifically *E. faecalis*, which are part of the natural flora of the intestinal tract and can cause serious infections in hospitalized patients. When the enterococci developed resistance to common antibiotics in the 1980s, physicians had responded by using vancomycin to treat them. Vancomycin-resistant *Enterococcus* (VRE) was first reported in 1986, and the *van*A gene soon spread throughout the species. Because enterococci readily exchange genetic information with other bacterial species, says Tenover, he and other experts assumed that it would soon pass *van*A and vancomycin resistance to MRSA. “Everybody was waiting for the shoe to drop,” says Tenover. In 2002 it did, when the Michigan Department of Community Health reported the first isolate of MRSA that had *van*A-mediated resistance to vancomycin. The patient was a 40-year-old diabetic who had recently been given an extended course of vancomycin for a foot ulcer.

Fortunately, vancomycin-resistant MRSA—now known as VRSA—has not developed into the nightmare researchers feared. Only nine isolates have been detected worldwide in 6 years, seven from the same region in Michigan, which suggests that *S. aureus*, unlike *Enterococcus*,

1940

1940
Penicillinase, an enzyme capable of destroying penicillin, identified in bacteria

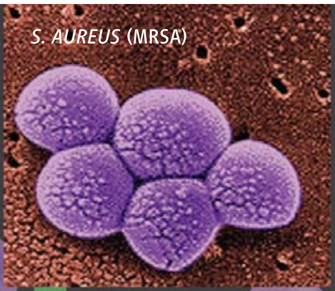
1942
First therapeutic use of penicillin

1943
Penicillin mass-produced

1945
More than 20% of *S. aureus* hospital isolates are penicillin-resistant as penicillinase begins to spread worldwide

1947
Streptomycin approved by FDA

1947
Streptomycin resistance observed



S. AUREUS (MRSA)

1952
Tetracycline approved by FDA

1956
Tetracycline resistance observed

1958
Vancomycin introduced, although rarely used until the mid-1980s

1959
Methicillin introduced

1961
Methicillin-resistant *S. aureus* (MRSA) observed

1964
Cephalothin, first antibiotic in the cephalosporin class, introduced

1966
Cephalothin resistance observed

1967
Gentamicin approved by FDA

1970
Gentamicin resistance observed

TABLE SOURCE: INFECTIOUS DISEASES SOCIETY OF AMERICA; (PHOTO) JIM BIDDLE/CDC; (TIMELINE SOURCE) C. T. BERGSTROM AND M. FELDGARDEN, *THE ECOLOGY AND EVOLUTION OF ANTIBIOTIC-RESISTANT BACTERIA*, IN S. STEARNS AND J. KOELLA, EDS., *EVOLUTION IN HEALTH AND DISEASE* (2008); KAREN BUSH/J&J

loses the ability to compete in the broader environment when it takes on the *van*A gene. That only one of these infections was life-threatening suggests that the vancomycin-resistant bug also loses its virulence.

A paltry pipeline

Although MRSA and other Gram-positive bacteria remain a major threat, a half-dozen new antibiotics have either just been approved or are in the pipeline that should work well against them—at least until the bugs evolve more resistance. This is not the case, however, for Gram-negative bacteria, such as *P. aeruginosa*, *A. baumannii*, and *K. pneumoniae*. These bacteria have both an inner and outer cell membrane, as opposed to the single cell membrane of Gram-positive bugs like MRSA. (The name comes from how these bacteria stain on a Gram stain test.) The pipeline for antibiotics against Gram-negative bacteria, says Bush of J&J, is limited to development programs in a few small companies; only one drug has made it through phase I clinical trials.

Prompted by the emergence of MRSA and VRE in the late 1980s, pharmaceutical companies focused their attention on Gram-positive bugs. Meanwhile, many Gram-negative bugs became resistant to virtually every known antibiotic, or at least every antibiotic that isn't toxic. "These organisms may well start to spread into the community," says Tenover, "and then we really will be in trouble. We have drugs to fall back on for *Staphylococcus*. But when you say, 'Where's the next anti-*Pseudomonas* drug?' I have to scratch my head."

One reason for the dearth of drug candidates is that Gram-negative bacteria are simply harder to kill. First, they have the extra cell membrane the drug has to penetrate. Then they have other defense mechanisms that Gram-positives lack, such as the ability to activate pumps or close down protein channels in the membranes that let these antibiotics in. "They can have three or four mechanisms working at once," says Case Western's Rice. "Even if you develop a new drug entirely, these bacteria may be just as likely to be resistant to new drugs as old ones. It's just really hard."

The problem has been exacerbated by the gradual exodus of pharmaceutical companies from antibiotic development—a trend that began in the 1980s and has accelerated since 2000, in large part because the market is iffy and the chances of success are slim. Of the 15 major pharmaceutical companies that once had flourishing antibiotic discovery programs, eight have left the field entirely, and two others have reduced their efforts significantly. That leaves only five—GlaxoSmithKline, Novartis, AstraZeneca, Merck, and Pfizer—that still have antibiotic discovery efforts commensurate with the size of the problem.

Estimated cases of hospital-acquired infections*

Antibiotic-Resistant Bacteria	Estimated Cases
Methicillin/*S. aureus*	102,000
Methicillin/CNS	130,000
Vancomycin/enterococci	26,000
Ceftazidime/*P. aeruginosa*	12,000
Ampicillin/*E. coli*	65,000
Imipenem/*P. aeruginosa*	16,000
Ceftazidime/*K. pneumoniae*	11,000

* Selected resistant bacteria, U.S., 2002

Even though the market for antibiotics is in the neighborhood of $25 billion a year, says Steve Projan, vice president of biological technologies at Wyeth Research in Cambridge, Massachusetts, other drugs, such as antidepressants or antihypertensives, offer a greater bang for the buck because they are often taken for years or decades rather than just a 7- to 14-day course. Resistance only compounds the problem: A drug that takes a decade to develop might be useful clinically for only a handful of years.

What's more, the better the antibiotic, the less health experts want to see it used to avoid the development of resistance. "It's probably the only area of medicine where a drug company can invest all this money to develop a drug, come up with a good one, and then the so-called thought leaders in the field, like myself, will tell people not to use it," says Rice. "We say it's such a good drug that we should save it."

As a result, virtually all the new antibiotics and all those in the pipeline for Gram-positive bacteria are second-generation drugs, that is, incremental improvements on existing classes. The one conspicuous exception—daptomycin, developed for *S. aureus* by the late Frank Tally at Cubist in Lexington, Massachusetts—was first identified 20 years ago by Eli Lilly and Co. and then shelved because it had toxicity problems at high doses.

To the surprise of many, the recent sequencing of more than 650 bacterial genomes has been a "dismal failure" when

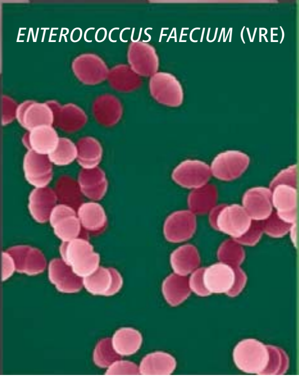
1976
Transferable penicillinase first observed in a gonococcus

1981
Cefotaxime approved by FDA

1983
Cefotaxime resistance observed

1983
First penicillin-resistant *Enterococcus* reported

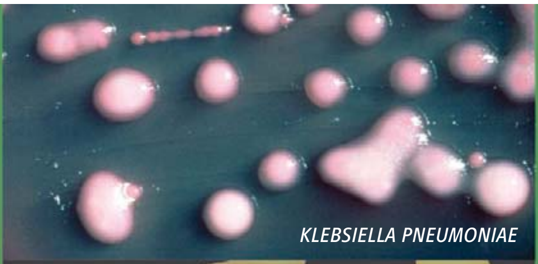
1987
Vancomycin-resistant *Enterococcus* (VRE) observed



ENTEROCOCCUS FAECIUM (VRE)

1987
First outbreak of *Klebsiella pneumoniae* resistant to third-generation cephalosporins

1996
S. aureus with intermediate resistance to vancomycin (VISA) reported



KLEBSIELLA PNEUMONIAE

1999
Community-acquired MRSA reported

2000
Linezolid, first antibiotic in the oxazolidinone class, approved by FDA

2001
Linezolid-resistant *S. aureus* and VRE observed

2002
S. aureus with complete resistance to vancomycin (VRSA) observed

2002

TABLE SOURCE: INFECTIOUS DISEASES SOCIETY OF AMERICA; (PHOTOS LEFT TO RIGHT) PHOTOTAKE INC./ALAMY; CDC; (ILLUSTRATIONS) N. KEVITIYAGALA/SCIENCE

it comes to drug development, says ICL's Enright. Although genome sequences were expected to yield a "treasure trove of new targets for entirely new classes of antibiotics," as David Pompliano and colleagues at GlaxoSmithKline in Collegeville, Pennsylvania, recently wrote, this simply hasn't panned out. At GlaxoSmithKline, Pompliano and his colleagues spent 7 years and more than $70 million evaluating more than 300 "canonical" bacterial genes that they thought were essential to the viability of the bacteria. The result was just five leads, a success rate, they estimated, that was four- to fivefold lower than for other areas of therapeutics.

Genomics is simply not a good paradigm for discovering new antibiotics, suggests Projan. The genetic approach assumes that a candidate drug can knock out a single gene in the bacterium to render it unfit for survival. But the drugs don't knock out a gene's activity entirely, he says; instead they modulate activity. "As we found out in oncology," says Projan, "sometimes leaving even 5% activity is enough for the tumor to grow. The same thing is true for bacteria." Projan

Collateral Damage: The Rise of Resistant *C. difficile*

In April 2002, Mark Miller, an infectious-disease specialist and microbiologist working at Jewish General Hospital in Montreal, Canada, began to suspect that he had an outbreak on his hands. He was used to dealing with the bacteria *Clostridium difficile*, which can cause severe diarrhea in debilitated patients and had been a common problem in hospitals for more than 30 years. But now the number of cases had started to climb, as did their severity. "One of the first indications that we knew we had a problem," says Miller, "was when one of the colorectal surgeons called me and said, 'I just took out my second colon in a month on a *C. difficile* patient.' When we started looking at the numbers, they were absolutely horrendous." At the peak of the outbreak, says Miller, there were 50 new cases of *C. difficile* diarrhea every month in their 600-bed hospital. "Of those, about one in six was dying or going for a colectomy. That's kind of staggering."

Resistance to antibiotics makes for bacteria that are harder to kill, but it can also bestow on a bacterial strain the advantage it needs to spread through the hospital environment and perhaps around the world—a kind of collateral damage in the escalating war between man and microbes. *C. difficile* is an unfortunate case in point. The bacterium has currently been linked to at least 5000 deaths a year in the United States; at the height of the Quebec epidemic it caused more than 7000 serious infections and 1200 deaths in a single year. In many hospitals, *C. difficile* constitutes a greater risk to patients than methicillin-resistant *Staphylococcus aureus* or any other bacteria.

The symptoms of a *C. difficile* infection range from mild diarrhea to severe colitis, and the elderly bear the brunt of the disease. One in four patients will have a recurrence or multiple recurrences. "It's a horrible problem," says Dale Gerding, an infectious-disease specialist at Hines Veterans Administration Hospital and Loyola University in Chicago, Illinois. Patients have to be treated almost constantly with oral vancomycin to prevent recurrences, he says.

C. difficile diarrhea first appeared in the medical literature in the 1970s, mistaken for a side effect of the antibiotic clindamycin. In 1978, physicians realized that the diarrhea was induced by toxins from clindamycin-resistant *C. difficile*, which had colonized the victim's colon after their normal gut flora had been decimated by the clindamycin treatment. *C. difficile* has remained a common hospital infection ever since because the bacteria produce heat-resistant spores that are exceedingly difficult to kill. "They're very resistant to detergents and cleaning agents," says Gerding. "Really, the only thing that destroys them is bleach or hydrogen peroxide."

Through the 1990s, however, *C. difficile* wasn't considered a major threat because the bacteria were susceptible to two antibiotics, vancomycin and metronidazole, the latter of which is inexpensive. As many as 40% of all hospitalized patients are colonized with *C. difficile*, but most tolerate it without symptoms. A series of hospital outbreaks in six U.S. states, beginning around 2000 and capped by the severe Quebec outbreak in a dozen hospitals, suggested that a new, hypervirulent strain of *C. difficile* was circulating.

Since then, the same offending strain has been identified in hospitals in 38 states and has also been linked to outbreaks in Western Europe. What sets it apart from its predecessors, say Gerding and Miller, is its high resistance to the newer fluoroquinolone antibiotics, such as levofloxacin and moxifloxacin. These antibiotics began to be used widely in the late 1980s, and usage has increased steadily ever since. Why this strain induces more severe disease—with a death rate among those infected of 10% compared with 1% percent in the 1980s—is still a mystery, but one possibility, says Gerding, is a mutation that enables the strain to produce more toxin.

Although Quebec hospitals have reduced the incidence of *C. difficile* infections by two-thirds since the height of the outbreak, through very tight isolation and control and rigorous "housekeeping," says Miller, they have yet to get back to the levels preoutbreak. "*C. difficile* in health-care facilities and hospitals is a very unforgiving organism," he says. It exploits "any lapse in isolation, in housekeeping, in hygiene—whatever it is—and it comes back with a vengeance."

–G.T.

The battle escalates. A hypervirulent strain of *C. difficile*, resistant to two of the newer, last-resort antibiotics, has triggered outbreaks across the United States and in Western Europe.

CREDITS (TOP TO BOTTOM): LOIS S. WIGGS/CDC; IAN NICHOLSON/PA PHOTOS/LANDOV

and others suggest that the only route to a new antibiotic—short of pure luck—will be through more fundamental research on the basic biology of the bacteria.

Cutting back

Barring the discovery of miracle antibiotics to which bacteria cannot evolve resistance—a "laughable" notion, says one expert—the only foreseeable route to curbing antibiotic resistance will be to rein in the use of antibiotics. One obvious way is to lower the risk of acquiring resistant bugs in the hospital. Countries that have mandated rigorous infection control in hospitals, such as Denmark, the Netherlands, and Finland, have been able to keep MRSA infection rates low. These infection-control procedures, however, go far beyond physicians and nurses wearing gloves and protective masks and washing their hands before and after patient contacts, essential as those are. These nations employ a technique known as "active detection and isolation," or "search and destroy," as it's called in the Netherlands. Patients considered at high risk of carrying MRSA and other antibiotic-resistant bugs are cultured when they're admitted to the hospital, and periodic cultures are taken of all patients, particularly those in high-risk wards. The greater the prevalence of pathogens and risk factors, the more frequent this surveillance. Patients who are infected or are carriers are isolated. Health-care workers who are colonized with resistant bacteria can be "decolonized," using skin washes and nasal ointments.

Whether U.S. hospitals should be required to implement active detection and isolation is a long-running controversy. Some specialists—led by University of Virginia epidemiologist Barry Farr, an expert on controlling VRE and MRSA—have argued that it's the only proven method to control hospital MRSA infections. Others have questioned the technique's cost-effectiveness and viability, particularly when rates of MRSA in the community are beginning to rival those in many hospitals.

Vaccines against antibiotic-resistant bacteria would also go a long way to reining in resistance, but only one such vaccine candidate, against *S. aureus*, has ever made it through phase III clinical trials: StaphVAX, licensed by Nabi Biopharmaceuticals. Although patients who received the vaccine had significantly lower rates of *S. aureus* infections at 40 weeks compared to controls, this apparent protection was lost at 54 weeks. A follow-up trial also failed to demonstrate efficacy. Several more vaccines are in development, including a new-generation StaphVAX. Even temporary protection could be useful, argue some experts, either for health-care workers, who could be vaccinated regularly, or for patients who are about to be hospitalized to undergo a procedure.

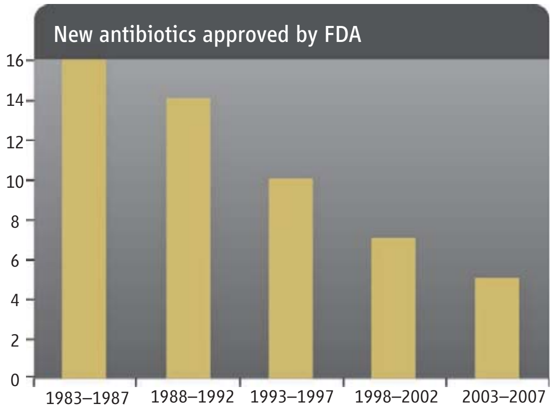
Ultimately, physicians will have to be persuaded to reduce their use of antibiotics, although that will be a hard sell. One step, for instance, would be to persuade physicians outside hospitals to treat only those patients who are truly infected. A 2001 study from the University of Colorado Health Sciences Center estimated that 55% of all antibiotics prescribed in the United States for upper respiratory infections were unnecessary. This is what Rice describes as the "get-a-little-sniffle-get-a-little-Levaquin" problem. "The patients want it," he says, and "the doctor wants to get the patient out of the office, and the quickest way to do it is to write a prescription." But the societal problem of antibiotic resistance should outweigh whatever personal peace of mind comes from the indiscriminate use of antibiotics, says Tenover (see sidebar on p. 360).

Similarly, many times physicians prescribe combination "broad-spectrum" antibiotics when a single "narrow-spectrum" antibiotic would do the trick. Understandably, says Rice, physicians are unwilling to wait to treat serious infections until the bug is cultured and they learn to which antibiotics it's still susceptible. But once the crisis is over, usually 1 to 3 days after starting therapy, physicians could switch their patients to the appropriate narrow-spectrum antibiotic.

What the field desperately needs, these experts say, are randomized, controlled trials to establish how long antibiotic therapy should be prescribed for different infections. The data are scarce, and misconceptions abound. The ubiquitous advice in the field—from physicians, patients, and even CDC—is that patients should continue the full course of antibiotics even after they feel better. Because antibiotics tend to have few side effects, physicians consider a longer course to be a no-lose proposition.

But from the perspective of preventing antibiotic resistance, says Rice, "this is totally wrong-headed." In patients with healthy immune systems, he explains, most antibiotics merely stun the bacteria sufficiently to make it easier for the host immune system to do its job. "You can take tetracycline until the cows come home," Rice says, and "all it does is stop most bacteria from growing. It doesn't kill them." Extending the course of the antibiotics unnecessarily increases the likelihood that the patient's normal flora will be inhibited to the point that bacteria resistant to the antibiotic will fill the void.

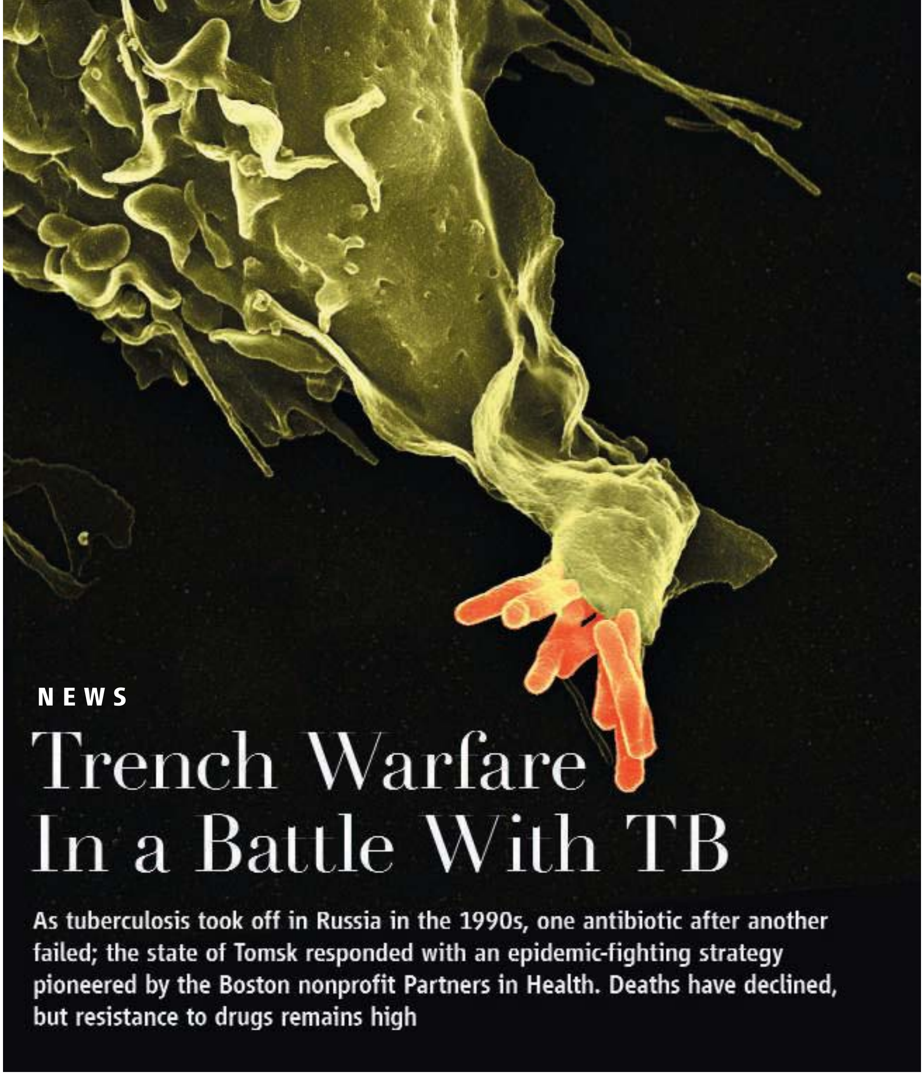
The few existing studies on the necessary length of therapy have suggested that it is often surprisingly short. Urinary tract infections in young women can be treated with 1- to 3-day courses of antibiotics. The Infectious Diseases Society of America recommends a 3-day treatment for traveler's diarrhea, while acknowledging that 1 day appears to be equally effective. Studies from the 1940s suggested that the "vast majority" of patients with pneumonia get better after 2 or 3 days, says Rice: "Somewhere along the line, that morphed into 7 days, 10 days, 21 days, with no real reason other than making the doctor more comfortable." In May, the U.S. National Institute of Allergy and Infectious Diseases responded to the expert demand and put out a request for proposals for clinical studies that would determine the optimal use of antibiotics, including the optimal duration of therapy. "I think most physicians would respond to compelling data from a well-done trial," says Rice.



SOURCE: INFECTIOUS DISEASES SOCIETY OF AMERICA

One beneficial side effect to curbing antibiotic use is that it may serve to rehabilitate those antibiotics that have lost their effectiveness. "Many of these were wonderful new drugs just 20 years ago, able to treat a wide variety of bugs, both inside and outside the hospital," says Rice. "Now we're at a point where some of them are next to useless, because they've been used for everything."

–GARY TAUBES



Tough bug. *Mycobacterium tuberculosis* (red) likes nothing better than to be ingested by a macrophage, its usual home.

NEWS

Trench Warfare In a Battle With TB

As tuberculosis took off in Russia in the 1990s, one antibiotic after another failed; the state of Tomsk responded with an epidemic-fighting strategy pioneered by the Boston nonprofit Partners in Health. Deaths have declined, but resistance to drugs remains high

TOMSK, RUSSIA—Six mornings a week, nurse Galina Kruchinina circles this Siberian city with a black medicine bag. She's hunting for people infected with tuberculosis—some of them rough customers—who don't show up at the clinic. She and a driver track them down and give them antibiotics. The search this morning leads down a rutted lane near the banks of the Tom River. They're looking for Oleg, 35, who will get three kinds of pills plus an injection, if they find him.

The car pulls up under some tall birch trees. Kruchinina disappears into a wooden building and brings down a sleepy-looking man. It's Oleg, who recently got out of prison, bringing *Mycobacterium tuberculosis* with him. Like many convicts, he has tattoos, including two stark blue ovals inked onto his eyelids that watch you when he blinks. Oleg takes his medicine in the car, accepting juice and food as part of the treatment from "Mother," as he calls Kruchinina.

Outreach. The Sputnik (companion) program deploys nurse Galina Kruchinina to find and treat TB patients.

If his infection isn't treated continuously for 6 months with a four-drug blitz, it could flourish and kill Oleg. A bigger worry for public health is that intermittent or weak therapy could spawn drug-resistant bacteria. Multidrug-resistant TB (MDR TB), as it's called, is on the rise here and across most of the globe, but luckily Oleg doesn't have it. If he keeps taking his antibiotics twice a day for several more months, he should be okay—which is why Kruchinina and her colleagues seek him out.

Kruchinina and her driver go next into a warren of rickety houses after Ivan, an alcoholic 21-year-old, and his older brother Ruslan, both infected. They've gone away for potatoes, someone says. The medical team doubts this; they track the men to a nearby house. Next they look for Liliana, who's been known to hide under the furniture; then Alexei, a brawny man in his 30s who lives with his mother. All the patients are difficult at first, says Kruchinina, "but they adapt." At 2 p.m., she gives her seat to another nurse, who completes the day's cycle.

This seek-and-treat program is called "Sputnik" ("companion" in Russian) by its creator, Partners In Health (PIH), a nonprofit linked to Harvard University's Brigham and Women's Hospital in Boston, Massachusetts. Sputnik workers are like family for the 28 people they follow. They reflect the PIH credo that no matter how poor or difficult the patient, the medic's duty is to deliver treatment.

PIH has joined with leaders of the Tomsk oblast, or state, and the Tomsk prison to deploy this aggressive strategy in a place where TB was raging out of control a decade ago. Sputnik is one component of Tomsk's effort, which in recent years has won millions of dollars in international aid. The result, says Harvard-based PIH physician Salmaan Keshavjee, is a program that would be "avant-garde, even in the United States."

PIH doesn't aim to hospitalize patients, as the Russian national TB system often does. Putting people in institutions tends to increase transmission of bacteria, PIH doctors argue. PIH focuses on home treatment or day clinics, with care by nurses or community workers rather than doctors, a method PIH pioneered in Haiti and Peru.

In 1999, PIH founders Paul Farmer and Jim Yong Kim began pushing for changes in international guidelines that once set aside MDR TB patients as untreatable. Farmer and Kim argued that the World Health Organization (WHO) should alter its policy to treat such patients, on both moral and economic grounds (*Science*, 10 August 2001, p. 1049). "If you don't treat MDR TB, it doesn't go away," says PIH doctor Michael Rich: "It only gets bigger and bigger." And the cost of treating each case of TB escalates.

PIH has put its ideas on the line here, making Tomsk a training center for dealing with TB in "resource-limited" jurisdictions. With 5 years of significant outside aid, the results are encouraging, but the statistics are still worrisome. The TB death rate is down; the cure rate is up; but the percentage of new infections that are drug-resistant is higher than in 2002, officially having risen to about 16%. And a new specter looms: extensively drug resistant, or XDR, TB.

CREDITS (TOP TO BOTTOM): SPL/PHOTO RESEARCHERS INC.; E. MARSHALL/*SCIENCE*

Where TB lives

M. tuberculosis is a tough organism, protected by a waxy lipid coat and an ability to hunker down and fend off immune and chemical attacks. It can travel in droplets expelled from the lungs—studies indicate it may live for several hours this way—and readily infects a new host. Lung infections are the most common, although other tissues can get infected. Poor nutrition increases risk, and some people can't handle the anti-TB drugs or may have an inherited vulnerability to TB. Coinfection with HIV makes TB more deadly, but Tomsk officials say they see relatively little of that.

TB is concentrated in a vast reservoir of the poor, particularly in prisons and hospitals. But everyone is in danger, says Edward Nardell, a PIH doctor, Tomsk consultant, and chief TB control officer for the state of Massachusetts. "The big risk factor is breathing."

Maria Rubina, a young patient at the main TB hospital in Tomsk, illustrates the point. She was diagnosed with TB when she was a student 3 years ago, at age 20. Since then, she has been living full-time in the Tomsk TB hospital, a big pink sanatorium built in 1937 across the river from town. No one knows for sure where she got the infection, but in retrospect, it seems she may have contracted it from a neighbor who died in her apartment complex (not in Tomsk). A postmortem showed that he had MDR TB.

Aivar Strelis, the top surgeon and a renowned TB expert at the hospital, has concluded that Rubina has XDR TB. This means that even the "second-line" products commonly used against MDR TB—the bulky injectables and less potent drugs with nastier side effects—cannot be given to her. A large part of her left lung has been removed. The infection is dormant, meaning that it has lapsed into the latent state of most TB infections, when bacteria are not ejected in sputum. WHO estimates there are a staggering 2 billion people in the world infected with this dormant type of TB. WHO also estimates there are 450,000 new cases of MDR TB each year.

Cases like Rubina's are the legacy of what Sergey Mishustin, the chief TB officer of Tomsk oblast, describes as a "collapse" of the health system after the end of the Soviet Union in the 1990s. He was formerly head TB physician in the Tomsk prison, part of the federal justice system. He recalls that there just wasn't enough money to buy medicines. Ventilation wasn't adequate in prisons or hospitals, and there was little effort to segregate MDR TB prisoners from so-called drug-susceptible cases. A short stay in prison could become a death sentence.

CREDITS (TOP TO BOTTOM): (PHOTO) E. MARSHALL/*SCIENCE*; (TABLES) WHO; SERGEY MISHUSTIN

"I was in horror," Mishustin says, to find that medics in one clinic were feeding TB patients whatever antibiotics they had at hand. Because there was no systematic testing for resistance, they were amplifying resistant organisms in patients already infected with them by wiping out only the competing bacteria. It's called an "epidemiological pump," an effective way to incubate lethal bacteria. In his own clinic, Mishustin says, he wouldn't allow antibiotics to be given this way. When people learned that he left thousands of dollars' worth of drugs on the shelf and threw them out on their expiration date, Mishustin says, investigators accused him of misconduct. But he says he was excused when TB experts came to his defense.

By 2000, the incidence of TB had increased among Russians to more than 107 cases per 100,000 people, according to WHO. For comparison, the U.S. rate is below 5 per 100,000 (see table). Since 2000, Tomsk has brought TB incidence to a "stable" state (102.7 per 100,000), Mishustin says.

TB burden in selected countries, 2006

	Incidence*	Mortality*
Russian Federation	107	17
China	99	15
India	168	28
South Africa	940	218
United States	4.6	0.2

*Rate per 100,000 population.

TB burden in Tomsk oblast

	2002	2007
Incidence*	114.2	102.7
Mortality*	18.6	12.2
Primary MDR TB, prison†	12.2%	18.7%
Primary MDR TB, civilian†	13.2%	14.1%

*Rate per 100,000 population. †Percent positive among those tested for resistance.

A British medical group, MERLIN (Medical Emergency Relief International), was the first, in 1994, to bring outside help to Tomsk. The group distributed microscopes around this Poland-sized oblast, parts of it unreachable by car, so that medics could examine sputum samples and quickly determine who was shedding infectious bacteria. Billionaire George Soros provided more support in the mid-1990s, then turned the operation over to PIH. Tomsk got a World Bank loan and, with PIH advice, won the first anti-TB grant (for about $11 million) from the Global Fund to Fight AIDS, Tuberculosis, and Malaria. Through the Global Fund, Eli Lilly and Co. in Indianapolis, Indiana, agreed to provide capreomycin and cycloserine at cost.

Last resort. Patients in the main tuberculosis hospital may have drug-resistant infections that require surgical removal of part of the lung, explains Aivars Strelis.

Prison and city clinicians say that the joint TB control system they've created here is unique. TB prison doctor Alexander Pushkarev took some Western visitors around to show off the improvements in June. Prisoners are now segregated into 14 wards and separated by medical status. MDR TB cases have a dorm and exercise area. All patients are registered in a 7000-record database, and by law, when they leave prison, medical details are shared with the Tomsk city operation overseen by chief doctor Vera Golubchikova. Her staff keeps tabs on each case and tries to get all of them cured—by Sputnik's team if need be. Both the prison hospital (1000 inmates) and the oblast TB hospital (370 beds, 1000 patients

Isolation. In Tomsk's 1000-inmate TB prison hospital, those with multidrug-resistant infections are kept in a separate ward.

per year) have new ventilation systems and—a recent addition—airtight cabinets to hold sputum samples while they're waiting to be tested. Basic improvements have made a difference. The prison hospital, which reported 60 TB deaths in one recent year, according to Pushkarev, has had none so far in 2008.

Making existing tools work

Even in well-funded Tomsk, resource limits are visible. The TB hospital's forbidding "department four" is not much more than a shelter for tough MDR TB cases, a way station on the final descent. In the larger "department three" building next door, MDR TB patients, including contagious ones, live four to a room.

The grim dorms and complex 2-year treatment regimens for MDR patients wouldn't be needed if doctors had a single pill they could give, say, once a day for a couple of weeks and knock out the mycobacterium. No such dream medicine is on the horizon. But several new prospects are in clinical development—for the first time in 4 decades—thanks to work by nonprofits such as the Global Alliance for TB Drug Development and a few companies (see sidebar). Perhaps one will be available in Tomsk in 5 to 10 years.

This is encouraging, but what excites Tomsk TB doctors is the near-term plan for rapid diagnostics. Tomsk oblast has bought a building for a dedicated diagnostic lab, says health department chief Albert Adamyan; by the end of the year, he hopes it will be using liquid media in drug-susceptibility testing.

Drug-susceptibility testing with solid culture media can take several months, and even that isn't done routinely across Russia. The liquid media should enable Tomsk clinics to get results in 3 weeks, officials say. Shorter diagnosis times make it possible to use medicine more efficiently, reduce MDR TB transmission, and attack the worst infections earlier—a critical factor in this battle, says PIH expert Rich.

"What we really need," Rich adds, "is a good, cheap, point-of-service test." Wealthier countries have access to polymerase chain reaction tests that monitor variable TB organism genes, signaling within 24 hours whether the strain is resistant to the first-line drugs isoniazid and rifampin. It's still expensive for Tomsk. But a new, gene-based test that costs $8 is being promised by WHO and the Foundation for Innovative New Diagnostics in Geneva, Switzerland.

For now, PIH leaders put most of their time and energy into teaching medical workers in places with a heavy TB burden how to use the cheapest susceptibility testing methods and how to deploy antibiotics already in the cabinet. Medics from 10 nations, for example, were in Tomsk in June for training by PIH and local physicians. Keshavjee chairs an international group called the Green Light Committee that works with drug companies to distribute high-quality TB medicines at cost to programs that meet its standards.

What Tomsk has done, says Neel Gandhi, the AIDS researcher at Albert Einstein College of Medicine in New York City who reported a famous outbreak of XDR TB in KwaZulu-Natal, South Africa, in 2006, "will create a center of excellence and a model" for others. "What works in Tomsk may not work" everywhere, partly because HIV is a bigger factor in Africa, he says; climate, government, and cultures differ. But Gandhi adds, "It would be wonderful to create a center of excellence in South Africa" like the one in Tomsk.

With creative drug cocktails and carefully staged protocols, it's been possible to extend the use of decades-old drugs that are losing much of their punch. But it requires vigilance and huge persistence—the kind provided by Sputnik's workers. That could prove a tall order elsewhere.

–ELIOT MARSHALL

Anti-TB Drugs: And Then There Were None

Two crucial antituberculosis drugs are valued today precisely because they didn't seem valuable a short time ago. Capreomycin and cycloserine were nobody's first choice when Eli Lilly and Co. in Indianapolis, Indiana, developed them as antibiotics in the 1950s, says Lilly Vice President Gail Cassell. Capreomycin must be injected; cycloserine causes psychosis in 1% or more of those who take it. Patients must be watched closely for central nervous system effects. But because the drugs were little used, bacteria were not widely exposed to them and didn't develop much resistance. Now they are among the best "second-line" treatments for people with multidrug-resistant TB (MDR TB).

Near-term TB drug prospects

Drug type	Testing stage
Gatifloxacin	Phase III
Moxifloxacin	Phase III
Diarylquinoline (TMC207)	Phase II
Nitroimidazole (OPC67683)	Phase IIA
Nitroimidazole (PA824)	Phase IIA
Pyrrole (LL3858)	Phase I
Diamine (SQ109)	Phase I

In 2001, however, Lilly decided to get out of antibiotics after a "very painful" review of its "limited resources," says Cassell. (Antibiotics are not big moneymakers.) In recent years, nonprofit groups have been working with Lilly and other companies to extend the life of the few effective drugs on a dwindling list. Lilly, for example, makes capreomycin and cycloserine available at cost to programs vetted by the so-called Green Light Committee of the Global Fund to Fight AIDS, Tuberculosis, and Malaria. The company is transferring manufacturing expertise to nonprofits as well. Other companies—such as Bayer, GlaxoSmithKline, Johnson & Johnson, and Novartis—are looking for ways to subsidize antibiotics for developing nations, which need them desperately.

Melvin Spigelman, clinical chief of the Global Alliance for TB Drug Development in New York City, recently drew up a list of the most promising TB drugs on the horizon; it's alarmingly short (see table). Some prospects, such as moxifloxacin, are simply refurbished versions of existing medicines. Even so, when pressed to say when one might be ready for wide use, experts speak of 3 to 4 years at best, not months.

–E.M.

CREDITS (TOP TO BOTTOM): E. MARSHALL/SCIENCE; MELVIN SPIGELMAN, JOURNAL OF INFECTIOUS DISEASES (2007)

PERSPECTIVE

Antibiotics and Antibiotic Resistance Genes in Natural Environments

José L. Martínez*

The large majority of antibiotics currently used for treating infections and the antibiotic resistance genes acquired by human pathogens each have an environmental origin. Recent work indicates that the function of these elements in their environmental reservoirs may be very distinct from the "weapon-shield" role they play in clinical settings. Changes in natural ecosystems, including the release of large amounts of antimicrobials, might alter the population dynamics of microorganisms, including selection of resistance, with consequences for human health that are difficult to predict.

The success of antibiotics for treating infections and, conversely, the risk that antibiotic resistance poses for human health has meant that research in this area has focused primarily on their role within clinical settings. In contrast, the function of antibiotics in natural (nonclinical) environments has received relatively little attention. Most antibiotics used for treating infections are produced by environmental microorganisms, meaning that genes for antibiotic resistance must also have emerged in nonclinical habitats (*1*). A better understanding of the ecological role for antibiotics and antibiotic resistance in nonclinical environments (Fig. 1) may eventually help to predict and counteract the emergence and future evolution of resistance (*2*).

Ecological Role for Antibiotics in Natural Environments

The successful finding of antibiotic-producing microorganisms in soil led to the idea that a primary ecological role for antibiotics is likely to be in inhibiting the growth of competitors (*3*). Because any chemical entity can be toxic at a sufficiently high concentration, it seems plausible that molecules selected by the pharmaceutical companies for their antibiotic properties at therapeutic concentrations would also have distinct functions at the lower concentrations probably encountered in nature (Fig. 2). For example, low concentrations of antibiotics trigger specific transcriptional changes that are independent of the bacterial stress response pathways (*4*) and may have beneficial consequences for the bacteria that modulate the interactions within microbial communities (*5*, *6*). Similarly, molecules formerly classified as delivering signals for intermicrobial communication have subsequently been found to possess demonstrable antibiotic activity (*5*). This capacity for signaling at low concentrations has further clinical implications because subinhibitory concentrations of antibiotics could favor bacterial virulence under certain conditions (*6*, *7*).

Why So Many Antibiotic Resistance Genes?

The huge number of antibiotic resistance genes found in the environment (*8*) raises the obvious question of why so many have evolved. Recent work has shown a pronounced breadth of utilization of antibiotics as a source of nutrients by bacteria, and it seems natural that this should have led to considerable levels of resistance (*9*). Equally, it seems clear why antibiotic-producing microorganisms should possess determinants to help them resist the action of the antibiotics they produce, but it is less obvious why bacteria that

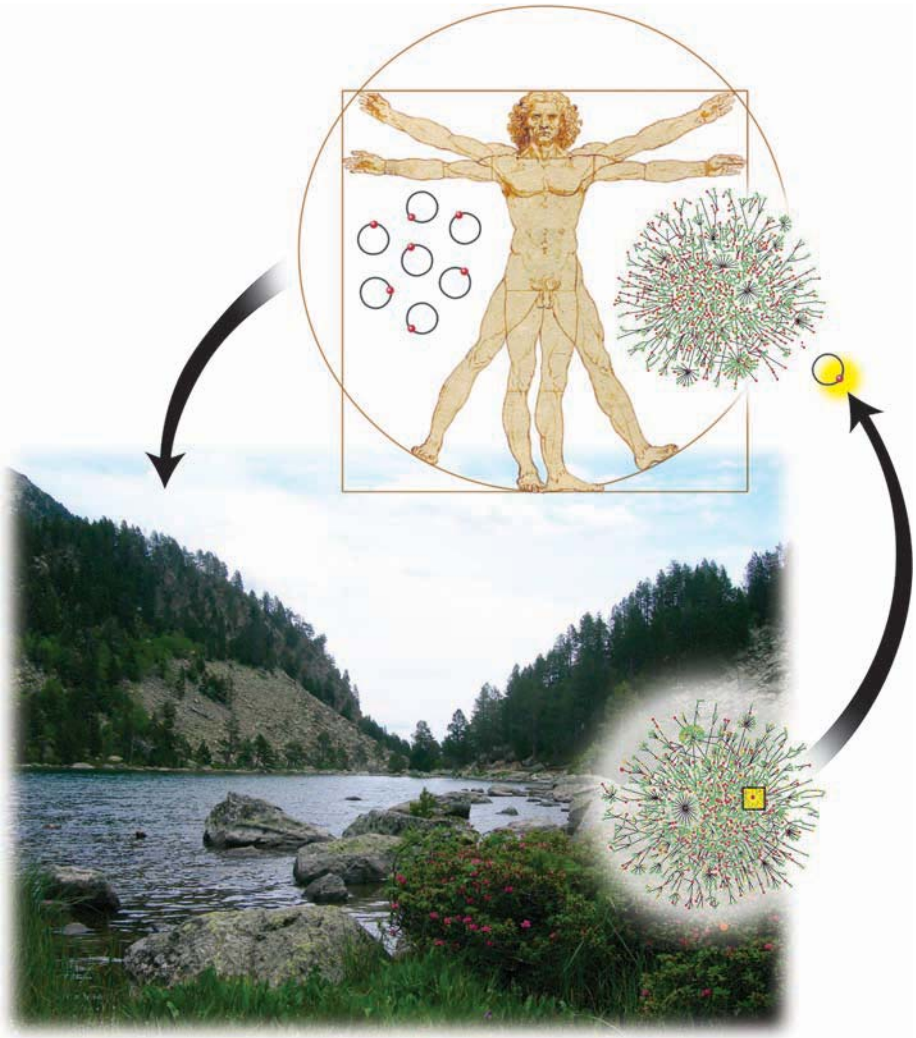


Fig. 1. Transfer of antibiotic resistance genes from natural clinical environments. Most antibiotic resistance genes acquired through horizontal gene transfer have been originated in environmental microorganisms. In their original host, these determinants (red dot) form part of integrated regulatory and metabolic networks. After their dissemination (circle), their integration in the novel metabolic networks would be difficult, and their only role will be resistance. The reintroduction of plasmid-encoded antibiotic resistance determinants in natural environments, together with the changes they suffer as the consequence of human activities, might be relevant for the future evolution and dissemination of antibiotic resistance determinants in bacterial pathogens.

Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnología (CSIC), Darwin 3, Campus UAM, Cantoblanco, 28049-Madrid, and CIBERESP, Spain.

*To whom correspondence should be addressed. E-mail: jlmtnez@cnb.csic.es

do not themselves produce antibiotics should also possess multiple resistance determinants (*10*). For example, multidrug resistance (MDR) efflux pumps are present in all organisms and can exist in large numbers within a single microorganism (*11*). Because their antibiotic resistance profiles overlap, it seems unlikely that all are required to resist antibiotics, and we now know that MDR elements are involved in other processes such as detoxification of metabolic intermediates, virulence, and signal trafficking, among other functions (*12–14*). What's more, a low-specificity enzyme might be classified as part of the antibiotic resistance machinery, even if it had evolved in the degradation or modification of metabolites (*15*). Thus, the dual nature of antibiotics as both signals and weapons can explain how genes, which may not have evolved their primary function in resistance, can nevertheless contribute to the protection from antibiotic threat.

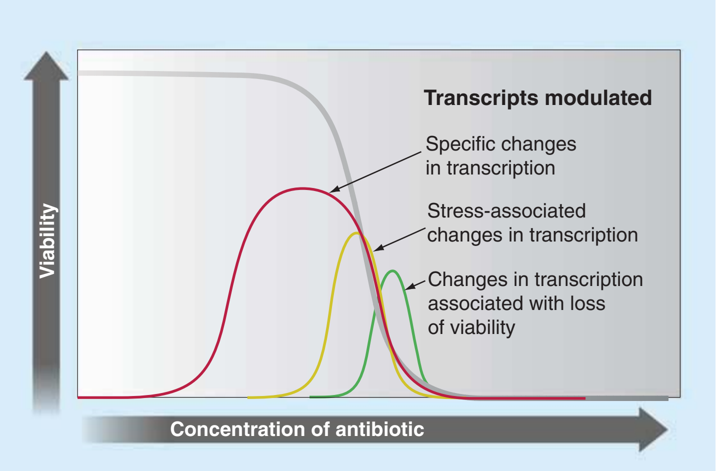


Fig. 2. Effect of antibiotics on bacterial gene expression. Antibiotics trigger concentration-dependent changes in bacterial transcription. At low concentrations, the expression of several genes changes, and this has been interpreted as a signaling effect. At higher concentrations, stress responses sum up to the changes in transcription and, at the highest concentrations, the changes in transcription probably reflect the loss of viability [modified from (*5*)].

Environmental Intrinsic Resistant Bacteria

Development of antibiotic resistance as the consequence of mutation (*16*) or horizontal gene transfer (*17*) is considered to have been driven by the relatively recent selective pressure of antibiotics used in therapeutic settings. Nevertheless, some bacterial species possess an intrinsically low susceptibility to antibiotics. Such bacteria have environmental origins in habitats that do not harbor a high antibiotic load, and those responsible for infections in hospitals (for instance, *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, or *Stenotrophomonas maltophilia*) are not themselves antibiotic producers. Free-living opportunistic pathogens often have large genomes that allow the colonization of diverse environments through metabolic versatility that helps to degrade and resist toxicity of compounds present in these ecosystems. This can include large numbers of biodegradative enzymes that cooperate in the modification and utilization of antibiotics as a food resource (*9*). Additionally, efflux pumps—originally involved in signal trafficking or resistance to toxic compounds produced by plants or rhizosphere-associated microbiota—can be used for effluxing antibiotics as well.

From Natural to Clinical Environments: A Functional Shift

In clinical environments, pathogenic and commensal bacteria are challenged with high concentrations of antibiotics. Thus, the function of the antibiotics in such ecosystems has been uniquely imposed by humans and is aimed solely at inhibiting bacterial proliferation. As a result, antibiotic resistance emerges to specifically overcome this inhibition. Drawing once again on the example of a protein able to function in modifying an antibiotic beyond its original function, it seems likely that this could serve as an antibiotic resistance determinant and evolve over shorter time frames in the clinic. For instance, *Providencia struartti* possesses an enzyme (2′-*N*-acetyltransferase**)** that modifies bacterial peptidoglycan, and the similarity of its substrate to gentamycin enables the enzyme to modify both its regular substrate and the antibiotic (*15*). Acquisition of the gene by another bacteria via plasmid transfer would place the encoded enzyme outside its normal biochemical context, and its unique function would then solely be acquired resistance to gentamycin (*18*).

The example described above addresses two characteristics that could allow functional shifting of antibiotic resistance determinants. Both refer to such elements as being "out of context." First, regulation of their expression in the pathogen is not equivalent to that in the original host. A key element in the establishment of metabolic networks is the fine-tuned regulation of the elements involved in response to external signals. Such regulation is often lost in the case of horizontally transferred antibiotic resistance genes, which are frequently expressed constitutively. For example, gene-capture units called integrons are involved in MDR and contain arrangements of antibiotic resistance genes, the expression of which is driven by a strong promoter (*19*). Second, the biochemical activity of a given protein encoded by a transferred gene is not inserted in the metabolic network where it evolved (Fig. 1).

Environmental Changes and the Evolution of Antibiotic Resistance

Natural environments represent reservoirs of antibiotic resistance genes, such that changes in these ecosystems might be relevant for the emergence of previously unknown resistance determinants in bacterial pathogens. It is surprising, therefore, that although there is a substantial concern over the potential effect of antibiotic resistance genes used for modifying organisms that can be released in the environment (*20*), the effect that changes to the environment may have on the population dynamics of bacteria and their antibiotic resistance genes has received much less attention (*1*). In this respect, it is worth noting that the antibiotic resistance genes most frequently used for genetic engineering are already widely disseminated among bacterial pathogens. In contrast, previously unrecognized antibiotic resistance genes that may emerge in the future already exist in many as yet ignored environmental organisms (*2*).

Whether anthropogenic changes of the environment might enrich the population of resistant bacteria and facilitate the transfer of resistance genes to human pathogens will be important to address in the future. The clearest example of this potential effect is contamination by antibiotics themselves. Antibiotics are currently used widely, not just for the treatment of human infections, but also in agriculture and animal/fish farming (*21*), with the possibility that high amounts of such compounds may find their way into natural habitats (*22*). An example of the effect of antibiotic contamination is that of the quinolone resistance gene *qnr*, which is present in the chromosomes of waterborne bacteria, where it has a so-far unknown function (*23*). After being integrated in plasmids, where it is constitutively expressed, *qnr* contributes to low-level resistance of its new bacterial host to quinolones (*24*). Recent work has shown that contamination of river waters by quinolones enriches for plasmid-encoded *qnr* genes present in waterborne bacteria, in such a way that may allow a first step in the transfer of this gene to human pathogens (*25*). Other types of contamination may also select for antibiotic resistance in nature. For instance, heavy metal pollution can select for antibiotic resistance (*26*), and stress conditions, as found in polluted environments, have the potential to increase recombination and horizontal gene transfer in a way that favors the dissemination of antibiotic resistance genes (*27*).

The presence of human commensal (and human-pathogenic) bacteria in the environment can be considered yet another form of contamination. Because any antibiotic resistance gene needs to coexist in the same environment as the human pathogen to which it may transfer, the increase in human population and the widespread lack of efficient wastewater treatment bring with them a risk of transfer of antibiotic resistance. Finally, it seems reasonable to speculate that a human-driven increase in the concentrations of antibiotics in natural ecosystems may not only influence antibiotic resistance, but also affect the broader microbial population dynamics in different natural environments.

Natural (nonclinical) habitats represent the main source of antibiotics and where antibiotic resistance has primarily evolved. The functional role these elements play in such environments is likely to be distinct from their "weapon/shield" function in clinical settings. In spite of the ecological relevance that antibiotics and resistance determinants have in nonclinical environments, there remains much to learn about the effect that human-driven changes of natural ecosystems may have on the evolution and dissemination of resistance in nature. Yet, the relevance this is likely to have for the future of human health is clear.

References and Notes

1. A. Alonso, P. Sanchez, J. L. Martínez, *Environ. Microbiol.* **3**, 1 (2001).
2. J. L. Martínez, F. Baquero, D. I. Andersson, *Nat. Rev. Microbiol.* **5**, 958 (2007).
3. S. A. Waksman, H. B. Woodruff, *J. Bacteriol.* **40**, 581 (1940).
4. E. B. Goh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17025 (2002).
5. A. Fajardo, J. L. Martínez, *Curr. Opin. Microbiol.* **11**, 161 (2008).
6. J. F. Linares, I. Gustafsson, F. Baquero, J. L. Martínez, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19484 (2006).
7. M. Gerber *et al.*, *J. Med. Microbiol.* **57**, 776 (2008).
8. V. M. D'Costa, K. M. McGrann, D. W. Hughes, G. D. Wright, *Science* **311**, 374 (2006).
9. G. Dantas, M. O. A. Sommer, R. D. Oluwasegun, G. M. Church, *Science* **320**, 100 (2008).
10. A. Fajardo *et al.*, *PLoS One* **3**, e1619 (2008).
11. J. Lubelski, W. N. Konings, A. J. Driessen, *Microbiol. Mol. Biol. Rev.* **71**, 463 (2007).
12. L. J. Piddock, *Nat. Rev. Microbiol.* **4**, 629 (2006).
13. J. F. Linares *et al.*, *J. Bacteriol.* **187**, 1384 (2005).
14. T. Kohler, C. van Delden, L. K. Curty, M. M. Hamzehpour, J. C. Pechere, *J. Bacteriol.* **183**, 5213 (2001).
15. D. R. Macinga, P. N. Rather, *Front. Biosci.* **4**, D132 (1999).
16. J. L. Martínez, F. Baquero, *Antimicrob. Agents Chemother.* **44**, 1771 (2000).
17. J. Davies, *Science* **264**, 375 (1994).
18. K. Franklin, A. J. Clarke, *Antimicrob. Agents Chemother.* **45**, 2238 (2001).
19. D. Mazel, *Nat. Rev. Microbiol.* **4**, 608 (2006).
20. S. Demaneche *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3957 (2008).
21. F. C. Cabello, *Environ. Microbiol.* **8**, 1137 (2006).
22. D. L. Smith, J. Dushoff, J. G. Morris, *PLoS Med.* **2**, e232 (2005).
23. L. Poirel, J. M. Rodriguez-Martinez, H. Mammeri, A. Liard, P. Nordmann, *Antimicrob. Agents Chemother.* **49**, 3523 (2005).
24. L. Martinez-Martinez, A. Pascual, G. A. Jacoby, *Lancet* **351**, 797 (1998).
25. V. Cattoir, L. Poirel, C. Aubert, C. J. Soussy, P. Nordmann, *Emerg. Infect. Dis.* **14**, 231 (2008).
26. A. Hernandez, R. P. Mellado, J. L. Martínez, *Appl. Environ. Microbiol.* **64**, 4317 (1998).
27. J. W. Beaber, B. Hochhut, M. K. Waldor, *Nature* **427**, 72 (2004).
28. J.L.M. is supported by grants LSHM-CT-2005-518152 and LSHM-CT-2005-018705 from the European Union and BIO2005-04278 from Ministerio de Educación y Ciencia.

10.1126/science.1159483

PERSPECTIVE

Outwitting Multidrug Resistance to Antifungals

Brian C. Monk and Andre Goffeau*

The economic cost of fungal infection and its mortality associated with multidrug resistance remain unacceptably high. Recent understanding of the transcriptional regulation of plasma membrane efflux pumps of modest specificity provides new avenues for the development of broad-spectrum fungicides. Together with improved diagnosis and indirect intervention via inhibition of the energy supply for drug efflux, we envisage multifunctional azole analogs that inhibit not only ergosterol biosynthesis and drug efflux-pump activity but also activation of the transcriptional machinery that induces drug efflux-pump expression.

Eight hundred million years of evolution have generated ~1.5 million fungal species that occupy many distinct ecological niches, yet only ~300 fungi cause disease in humans (*1*). The identification of antifungals that act specifically against these pathogens is a particular challenge because of fungal diversity, individualized pathways for infection, and fungal use of multiple mechanisms that circumvent exogenous toxins. These highly regulated mechanisms include innate resistance to specific antifungal drugs, formation of biofilms, selection of spontaneous mutations that increase expression or decrease susceptibility of the drug target (*2*), stress-related tolerance that enhances short-term survival (*3*, *4*), modification of chromosomal ploidy (*5*), and overexpression of multidrug efflux pumps (*6*). Fortunately, compared with infections caused by drug-resistant bacteria, those caused by resistant fungal pathogens and their spread to other patients occur relatively infrequently. However, the economic cost of fungal infection and its associated mortality, especially in debilitated and high-investment patients, remain unacceptably high.

A Clinical Perspective

The most prominent fungal pathogens affecting humans include *Aspergillus fumigatus*, *Candida albicans*, *C. glabrata*, *C. parasilosis*, *C. tropicalis*, *C. krusei*, and *Cryptococcus neoformans* (*7*). Although the skin, mucosal surfaces, and immune system usually provide robust defenses, weakened immunodefenses dramatically increase susceptibility to debilitating and life-threatening opportunistic fungal infections. Fungal infections are normally treated with a modest repertoire of drugs derived from five antifungal classes that target DNA and RNA synthesis, ergosterol, the ergosterol biosynthetic pathway, or the biosynthesis of the cell-wall component 1,3-β-D-glucan (Table 1). Unfortunately, the prophylactic use of fungistatic azoles such as fluconazole has been associated with an increased frequency of innate or acquired drug resistance in clinical isolates and the selection of non-*albicans Candida*, non-*fumigatus Aspergillus*, opportunistic yeastlike fungi, zygomycetes, and hyaline molds. Despite the fact that broader-spectrum third-generation azole drugs and the more expensive echinocandin class of antifungals prevent an increased proportion of life-threatening infections, *Candida* species remain the fourth most common cause of hospital-acquired bloodstream infection and kill 40% of those patients, whereas disseminated *Aspergillus* infections kill up to 80% of affected patients.

Mechanisms of Multidrug Resistance

Because of its economic and clinical impact, a focus on multidrug resistance rather than resistance to specific antifungals in pathogenic fungi is timely. Multidrug resistance, called pleiotropic drug resistance (PDR) in *Saccharomyces cerevisiae*, is an ancient phenomenon that preceded the modern use of antifungals (*8*). The adenosine triphosphate (ATP)–binding cassette (ABC) and major facilitator superfamily (MFS) transporter families responsible for multidrug resistance operate in all fungi. We distinguish among the transporters that belong to different species by using the prefix Sc for *S. cerevisiae*, Cg for *C. glabrata*, or Ca for *C. albicans*.

Saccharomyces cerevisiae. PDR in *S. cerevisiae* is the best-understood multidrug resistance mechanism in fungi. Point mutations conferring resistance to chemically diverse drugs (including azoles) have been mapped in genes encoding the zinc-

---

Department of Oral Sciences, Faculty of Dentistry, University of Otago, Post Office Box 647, Dunedin, New Zealand; and Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-Neuve, 1348, Belgium.

*To whom correspondence should be addressed. E-mail: andre.goffeau@uclouvain.be

finger transcription factors ScPdr1p or ScPdr3p (*9*, *10*). These gain-of-function mutations activate over 20 target genes, the major ones being either ATP-driven (ABC transporter genes *ScPDR5*, *ScSNQ2*, and *ScYOR1*) or proton motive force–driven (MFS transporter genes *ScTPO1* and *ScFLR1*) efflux pumps (*11*, *12*). Resistance to a wide spectrum of drugs is conferred via the activation of efflux-pump gene expression, which involves the binding of Pdr1p/Pdr3p to the consensus binding element PDRE (*13*, *14*). Mechanisms regulating PDR in *S. cerevisiae* include mutation of *PDR1/3*, plasma membrane sphingolipid homeostasis, ScPdr3p autoregulation, ScPdr3p-specific activation due to loss of mitochondrial respiration, chaperone-specific differential regulation of ScPdr1p and ScPdr3p (*15*), and ScPdr1p-dependent compensatory expression of efflux pumps (*16*). Yeast cells incubated with antifungals and other drugs transiently activate ScPdr1p/Pdr3p (*16*). Drugs such as itraconazole and progesterone bind to a 250–amino acid hydrophobic xenobiotic binding domain (XBD) of ScPdr1p/Pdr3p, enabling a specific association with the KIX domain of the Gal11p subunit of the mediator complex that recruits RNA polymerase II for expression of the ScPdr1p/Pdr3p-controlled genes (*17*) (Fig. 1A). Other transcription factors such as Yrr1p, Stb5p, Rdr1p, Yrm1p, and Yap1p also contribute to the expression of the various efflux transporter genes (*17*, *18*).

Pathogenic fungi. The human pathogen *C. glabrata* uses the transcription factor CgPdr1p to control expression of the ABC multidrug efflux pumps CgCdr1p and CgCdr2p through mechanisms very similar to those of its close relative *S. cerevisiae*. The pumps are induced by treatment with diverse drugs and are highly expressed in respiration-defective mutants. Antifungal binding to the CgPdr1p XBD induces multidrug resistance via a KIX domain from the *C. glabrata* mediator complex (*19*). Mutants overexpressing CgPdr1p coordinately regulate 11 genes homologous to ScPdr1/ScPdr3p targets (*20*). These similarities support the use of *S. cerevisiae* in developing tools that are directly applicable to antifungal resistance in *C. glabrata*.

Multiple azole resistance stemming from long-term prophylaxis is frequently found in clinical isolates of the more distant pathogen *C. albicans* (*21*). The resistance mechanisms of this diploid species are complex and include mutations in single genes, loss of heterozygosity, chromosomal rearrangements, and selective segregation of chromosomal fragments (*22*). About 85% of fluconazole-resistant clinical isolates show multidrug resistance due to overexpression of the ABC transporters CaCdr1p and CaCdr2p (homologs of the *S. cerevisiae* ScPdr5p) and the major facilitator superfamily (MFS) pump CaMdr1p (homolog of ScFlr1p). The efflux functions of these transporters can be cloned in *S. cerevisiae* (*23*, *24*). Expression of the CaCdr1p and CaCdr2p pumps is controlled by the transcription factor CaTac1p (*25*), which shares about 20% identity with ScPdr1/ScPdr3p. CaTac1p and ScPdr1p/ScPdr3p recognize substantially different PDREs (*25*), and CaTac1p causes more focused transcription than ScPdr1p/Pdr3p (*26*, *27*). High doses of the female steroid hormone progesterone transiently upregulate, via steroid-specific PDREs, the same core of ABC transporters induced by antifungal intervention or gain-of-function mutations in the transcription factors (*26*, *28*). Fluconazole-resistant clinical isolates often constitutively overexpress the MFS transporter CaMdr1p, either by itself or in combination with the azole target CaErg11p and/or the CaCdr1p and CaCdr2p ABC pumps. Although the MFS transporter CaMdr1p seems more efficient than its homolog ScFlr1p, it is overexpression of ABC transporters that confers clinically important, high-level azole resistance.

Preliminary data on non-*albicans Candida* species, *Cryptococcus neoformans*, and *A. fumigatus* (*29*) suggest that various resistance phenomena identified in *C. albicans* may operate in these pathogenic fungi and that PDR-related transcriptional mechanisms may contribute to their multidrug resistance.

Prospects

The long-awaited structural resolution of antifungal binding sites in the azole target ScErg11p as well as in the drug efflux pumps related to ScPdr5p would undoubtedly provide insight into multidrug resistance and guide strategies for impairing their activities. Meanwhile, and despite molecular mechanisms of differing complexity contributing to multidrug resistance in pathogenic fungi, a newly detected Achilles heel may be the transcriptional control of the antifungal efflux pumps. Of particular interest is the discovery that *PDR* transcriptional activators bind substrates of the efflux pumps they induce. We therefore may anticipate the development of novel multifunctional azole analogs. Erg11p would still be their primary target. Inclusion of a novel substituent would then enable inhibition of XBD-dependent coupling of Pdr1p/Pdr3p with its cognate mediator complex plus physical blockade of efflux via PDR transporters (Fig. 1B). The structures of itraconazole and fluconazole suggest that a fluconazole-like scaffold could be modified to antagonize not only Erg11p but also the transcriptional XBD and the active site from efflux pumps. Similarly, the dependence of the transient steroid response on interactions with the XBD domain of CgPdr1p, and possibly CaTac1p, indicates that a steroid hormone antagonist could increase the potency of azoles used against vaginal infections. Functional overexpression of Erg11p and both MFS and ABC drug efflux pumps from pathogenic fungi has been accomplished with an activated *PDR5* promoter in a *S. cerevisiae* host whose major *PDR* genes had been deleted (*24*). This approach has allowed the assessment of innate and overexpression-related resistance to antifungals and the discovery of efflux-pump inhibitors. Similarly, ScPdr1p-regulated overexpression of functional ScPdr5p-related pump homologs is expected to provide screens for the identification of the novel broad-spectrum azoles or narrower-spectrum steroid antagonists hypothesized above (*30*). By minimizing pump expression, drug pump activity, and the opportunity for stress responses, these drugs should transform the fungistatic azoles into potent fungicides.

A complementary strategy is the identification of new targets whose dysfunction kills fungi rapidly, thus avoiding the emergence of both drug

Table 1. Effects of antifungals.

Antifungal class	Drugs used in the clinic	Primary molecular target (mode of action)
Fluorinated pyrimidine analogs	5-Flucytosine	RNA and DNA biosynthesis (misincorporation of 5-fluorouridine)
Polyenes	Nystatin Amphotericin B	Cell-membrane ergosterol (increased permeability of plasma membrane and oxidative damage)
Allylamines and thiocarbamates	Terbinafine (for dermatophytes)	Ergosterol biosynthesis, squalene epoxidase, Erg1p (fungistatic inhibition of ergosterol biosynthesis)
Azoles and triazoles	Azoles Miconazole Triazoles Fluconazole Itraconazole Posaconazole Voriconazole Ravuconazole	Ergosterol biosynthesis, lanosterol 14α-demethylase, Erg11p (fungistatic inhibition of ergosterol biosynthesis and accumulation of toxic sterol intermediates)
Echinocandins	Caspofungin Micofungin Anidulofungin	Cell-wall biosynthesis, 1,3-β-D-glucan synthase, Fks1/2p (fungicidal or fungistatic inhibition of 1,3 β-D-glucan biosynthesis)

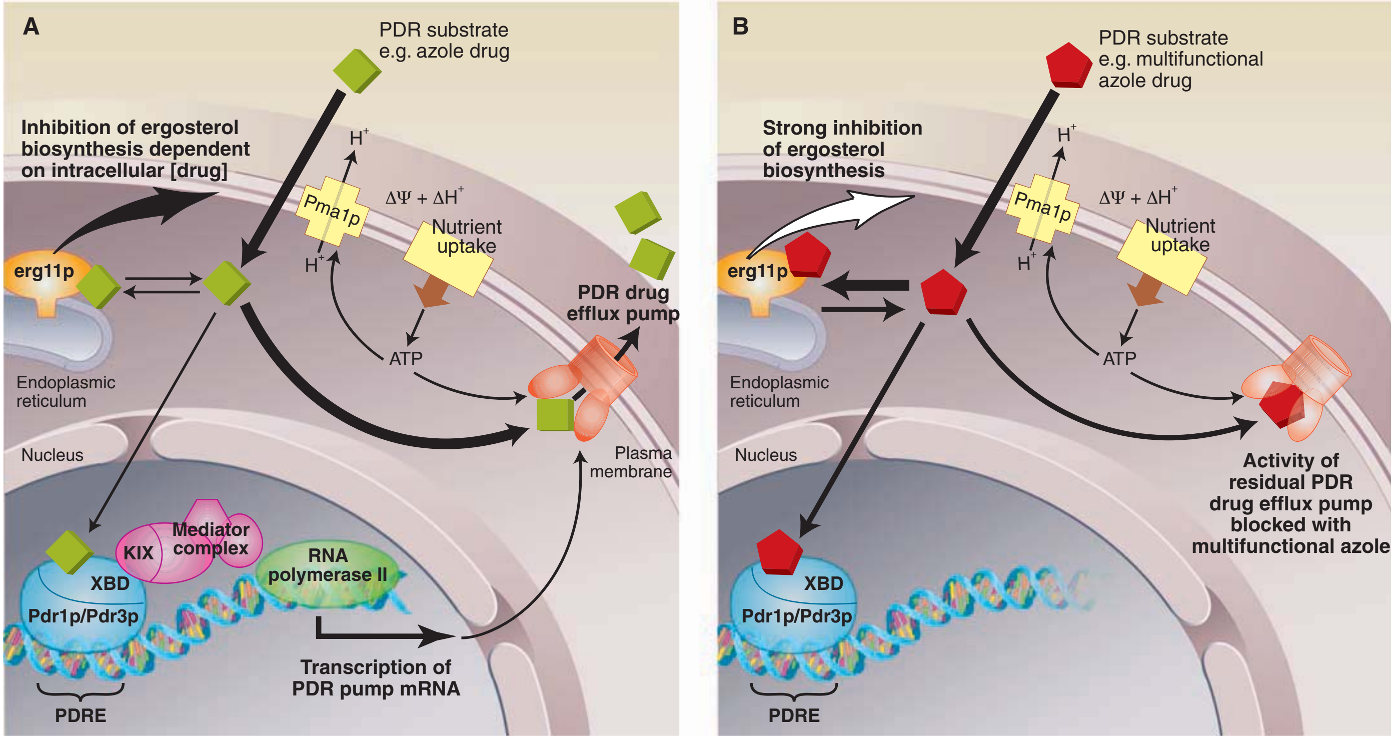


Fig. 1. Circumventing the PDR pathway in fungi. (**A**) In cells with clinically important resistance to azole drugs, high-level transcription of PDR efflux-pump genes involves the recruitment of RNA polymerase II, which depends on a drug-induced interaction between the ScPdr1p/Pdr3p and mediator complexes. The efflux pumps reduce the intracellular concentration of the drug below that required to inhibit the azole target Erg11p, allowing normal cell growth. (**B**) Binding of a multifunctional azole to the XBD domain of ScPdr1p/Pdr3p blocks expression of the drug pumps responsible for multidrug efflux and inhibits drug efflux by occupying a binding site in residual efflux pumps. The intracellular concentration of the drug is thus sufficient to block ergosterol biosynthesis in the endoplasmic reticulum. Reduced ergosterol content of membranes, production of toxic methylated sterols, and oxidative damage kill the fungal cell. Alternatively, other antifungals directly inhibit the electrogenic plasma membrane H$^+$-ATPase Pma1p, preventing the uptake of nutrients driven by the plasma membrane electrochemical gradient. The cells die rapidly because of a limited cellular energy supply and a loss of ion balance. Partial inhibition of Pma1p activity compromises the activity of both MFS and ABC transporters and increases the potency of azole drugs.

tolerance and efflux-mediated resistance. About 250 genes deemed essential in *S. cerevisiae* encode products that are at least 40% conserved across a broad range of fungi, including the fungal pathogens *C. glabrata*, *C. albicans*, *C. neoformans*, and *A. fumigatus* (*31*). Only about 50 of these gene products show less than 40% homology with human proteins. One of these is the plasma membrane proton pump (Pma1p), which generates the electrochemical gradient that fungi require for ion balance, nutrient uptake, and energy production. Pma1p inhibitors are fungicidal, indirectly block the activity of both ABC and MFS drug efflux pumps (*32*), and substantial resistance to them has yet to be detected.

Finally, diagnosis of disseminated fungal infections is too slow because conventional identification requires phenotypic examination of colonies grown for at least 48 hours on selective medium. The identification of drug resistance often requires a further step. Polymerase chain reaction amplification of ribosomal RNA intervening transcribed sequences followed by DNA pyrosequencing should halve the time needed for species-level fungal identification (*33*, *34*). Translation of this technology into the clinic will allow the early identification of fungal species, including innately resistant species or those susceptible to the development of multidrug resistance. The application of appropriate prophylaxis with existing and novel antifungals and of ongoing surveillance will save many lives.

References and Notes

1. L. H. Taylor, S. M. Latham, M. E. Woolhouse, *Philos. Trans. R. Soc. London Ser. B* **356**, 983 (2001).
2. D. Sanglard, J. Billie, in *Candidia and Candidiasis*, R. Calderone, Ed. (American Society for Microbiology Press, Washington, DC, 2002), pp. 349–383.
3. R. D. Cannon *et al.*, *Microbiology* **153**, 3211 (2007).
4. L. E. Cowen, *Nat. Rev. Microbiol.* **6**, 187 (2008).
5. A. Selmecki, M. Gerami-Nejad, C. Paulson, A. Forche, J. Berman, *Mol. Microbiol.* **68**, 624 (2008).
6. E. Balzi, M. Wang, S. Leterme, L. Van Dyck, A. Goffeau, *J. Biol. Chem.* **269**, 2206 (1994).
7. M. A. Pfaller, D. J. Diekema, *Clin. Microbiol. Rev.* **20**, 133 (2007).
8. Y. Gbelska, J. J. Krijger, K. D. Breunig, *FEMS Yeast Res.* **6**, 345 (2006).
9. E. Carvajal, H. B. van den Hazel, A. Cybularz-Kolaczkowska, E. Balzi, A. Goffeau, *Mol. Gen. Genet.* **256**, 406 (1997).
10. A. Nourani, D. Papajova, A. Delahodde, C. Jacq, J. Subik, *Mol. Gen. Genet.* **256**, 397 (1997).
11. J. DeRisi *et al.*, *FEBS Lett.* **470**, 156 (2000).
12. B. Rogers *et al.*, *J. Mol. Microbiol. Biotechnol.* **3**, 207 (2001).
13. D. J. Katzmann, T. C. Hallstrom, Y. Mahe, W. S. Moye-Rowley, *J. Biol. Chem.* **271**, 23049 (1996).
14. Y. M. Mamnun, R. Pandjaitan, Y. Mahe, A. Delahodde, K. Kuchler, *Mol. Microbiol.* **46**, 1429 (2002).
15. K. Gulshan, W. S. Moye-Rowley, *Eukaryot. Cell* **6**, 1933 (2007).
16. A. Kolaczkowska, M. Kolaczkowski, A. Goffeau, W. S. Moye-Rowley, *FEBS Lett.* **582**, 977 (2008).
17. B. Akache, S. MacPherson, M. A. Sylvain, B. Turcotte, *J. Biol. Chem.* **279**, 27855 (2004).
18. S. Le Crom *et al.*, *Mol. Cell. Biol.* **22**, 2642 (2002).
19. J. K. Thakur *et al.*, *Nature* **452**, 604 (2008).
20. J. P. Vermitsky *et al.*, *Mol. Microbiol.* **61**, 704 (2006).
21. T. C. White, K. A. Marr, R. A. Bowden, *Clin. Microbiol. Rev.* **11**, 382 (1998).
22. A. Coste *et al.*, *Eukaryot. Cell* **6**, 1889 (2007).
23. R. Prasad, P. De Wergifosse, A. Goffeau, E. Balzi, *Curr. Genet.* **27**, 320 (1995).
24. E. Lamping *et al.*, *Eukaryot. Cell* **6**, 1150 (2007).
25. A. T. Coste, M. Karababa, F. Ischer, J. Bille, D. Sanglard, *Eukaryot. Cell* **3**, 1639 (2004).
26. D. Banerjee *et al.*, *Eukaryot. Cell* **7**, 68 (2008).
27. T. T. Liu *et al.*, *Eukaryot. Cell* **6**, 2122 (2007).
28. B. Larsen, S. Anderson, A. Brockman, M. Essmann, M. Schmidt, *Yeast* **23**, 795 (2006).
29. A. M. Nascimento *et al.*, *Antimicrob. Agents Chemother.* **47**, 1719 (2003).
30. J. Cernicka *et al.*, *Int. J. Antimicrob. Agents* **29**, 170 (2007).
31. M. Liu *et al.*, *Eukaryot. Cell* **5**, 638 (2006).
32. B. C. Monk *et al.*, *Antimicrob. Agents Chemother.* **49**, 57 (2005).
33. A. M. Borman, C. J. Linton, S. J. Miles, E. M. Johnson, *J. Antimicrob. Chemother.* **61** (suppl. 1), i7 (2008).
34. B. L. Boyanton Jr., R. A. Luna, L. R. Fasciano, K. G. Menne, J. Versalovic, *Arch. Pathol. Lab. Med.* **132**, 667 (2008).
35. Supported by NIH grant DE016885.

10.1126/science.1159746

Ice Scour Disturbance in Antarctic Waters

Dan A. Smale, Kirsty M. Brown, David K. A. Barnes,* Keiron P. P. Fraser, Andrew Clarke

Disturbance has long been recognized by ecologists as a key factor regulating diversity on moderate to regional scales (*1*). The problem, however, has been experimentally determining frequency and intensity of the disturbance, particularly in the marine realm. Ice scour is a key structural force acting on benthic communities at high latitudes. It affects one-third of the world's coastlines, yet relatively little is known about the frequencies and the ecological effects of ice scouring compared with other natural disturbances. At small spatial scales, ice scouring may cause high faunal mortality, skewed population structures, and a dominance of mobile secondary consumers (*2*). At regional spatial scales, ice scouring may promote biodiversity by increasing habitat heterogeneity and driving the coexistence of different successional stages (*3*).

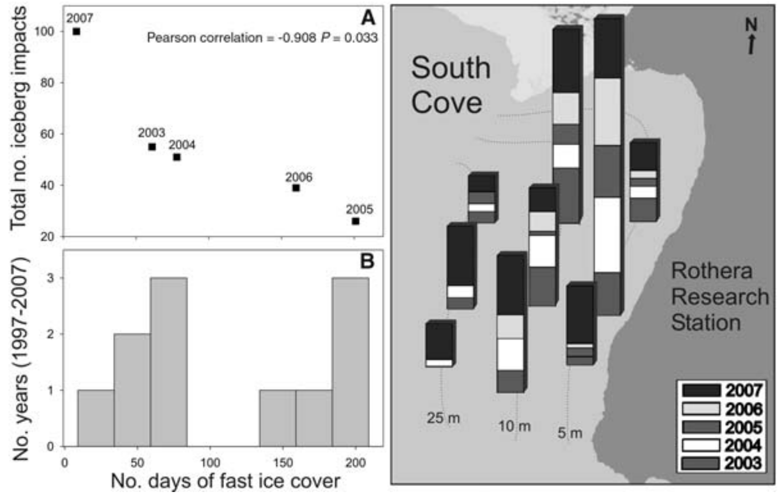


Fig. 1. (**Left**) Frequency of iceberg impacts decreases with increasing time that the sea surface is frozen (fast ice). The points are total number of damaged markers plotted against the number of days of fast ice (defined as 100% coverage of local area) for each year of the study (**A**). Histogram of the number of fast ice days recorded annually at the study site, since detailed records began in 1997 (**B**). (**Right**) Stacked bars indicate the number of damaged markers recorded at each disturbance grid for each year. The base of the bars indicates the approximate position of the grid within the study site, whereas contour lines indicate depth.

We measured the frequency of ice scouring at a shallow water site (South Cove, Rothera Point, Adelaide Island) on the West Antarctic Peninsula (WAP) for 5 years with an array of unique experimental markers. The markers were laid in grid formations, in triplicate, at 5-, 10-, and 25-m depths. Each January from 2004 to 2008, scuba divers surveyed the markers to determine the frequency of ice disturbance events at the site (see supporting online material for details).

We recorded highly significant variation between years and depths in the number of affected markers at each grid (two-way analysis of variance on untransformed data: for year, $F_{4,30}$ = 3.67 and P = 0.015; for depth, $F_{2,30}$ = 4.52 and P = 0.019; and for interaction, $F_{8,30}$ = 0.35 and P = 0.936). The total number of damaged markers (all depths) per year ranged from 26 (11.5% of all markers) in 2005 to 100 (44.4%) in 2007. In addition, the frequency of ice scouring at South Cove was strongly negatively correlated with the duration of fast ice coverage (i.e., the number of days per year when the immediate area was entirely covered with attached sea ice) (Fig. 1).

The proportion of damaged markers from each grid ranged from 0 to 76% per year. Overall, the grids laid at 25-m depth were less disturbed by icebergs than those at 10-m and 5-m depths, but we observed considerable variability between replicate grids at each depth increment (Fig. 1). For example, the most-disturbed grid at 5-m depth was affected by ice four times more frequently than the least-disturbed grid, despite the grids being only ~15 m apart. Also, for each grid location we recorded high variability in the number of iceberg impacts between years (Fig. 1).

Many factors will influence the probability of an iceberg impacting a given patch of seabed. These include depth, seabed topography, proximity to a source of icebergs, and local topographic influences such as the prevailing wind direction, tidal regime, and local and regional oceanography. Also important will be the local decay rate of icebergs and winter fast ice dynamics. Despite these sources of variability, we observed a very strong relationship between the frequency of ice scour and just one variable, the duration of winter fast ice. The functional basis of the relationship is straightforward: Once frozen into winter fast ice, icebergs are locked into position and no longer able to move or scour.

Although links between the duration of winter fast ice and ice scour frequency have been alluded to before, an empirical test of the relationship had not been attempted previously. Following the recent rapid warming of the WAP (*4*), significant changes in the extent of winter sea ice have been recorded in the region. Because our study was conducted at very shallow depths at a single site, we must exercise great caution when making predictions at the larger scale of the shelf ecosystem. Even so, our results would suggest that continuing reductions in nearshore winter sea ice duration would lead to an increased frequency of ice scouring to benthic communities inhabiting the shelf surrounding the WAP (to depths of 500 m, the maximum keel depth of modern icebergs). Furthermore, most maritime glaciers in the region have retreated over recent decades (*5*), which intuitively seems likely to have increased ice loading into coastal waters and thus the number of icebergs that currently float around the WAP and disturb the benthos.

Ice disturbance plays a key role in determining marine biodiversity at high latitudes. It seems likely that (i) ice scouring will intensify in Antarctic regions experiencing warming, probably over ecological time scales, and (ii) the benthic structure and diversity in these regions will be considerably affected by changing disturbance pressures (*6*). Conversely, on longer time scales (centuries) we envisage a drastic reduction of ice disturbance, as glaciers of the Antarctic Peninsula retreat past their grounding lines and the amount of free-floating sea ice (which scours the very shallows) decreases.

References and Notes

1. J. H. Connell, *Science* **199**, 1302 (1978).
2. D. K. A. Barnes, K. E. Conlan, *Philos. Trans. R. Soc. London Ser. B* **362**, 11 (2007).
3. J. Gutt, D. Piepenburg, *Mar. Ecol. Prog. Ser.* **253**, 77 (2003).
4. D. G. Vaughan *et al*., *Clim. Change* **60**, 243 (2003).
5. A. J. Cook, A. J. Fox, D. G. Vaughan, J. G. Ferrigno, *Science* **308**, 541 (2005).
6. D. A. Smale, D. K. A. Barnes, *Ecography* **31**, 289 (2008).
7. This report is dedicated to Kirsty Brown, who died at Rothera in 2003 while working on this project. We thank all members of the various marine teams that have worked at Rothera Research Station in the past decade. In particular, we thank dive officers J. Withers, M. Brown, and K. Murray; marine assistants A. Miller, P. Mann, and A. Massey; and L. Peck and S. Morley. The UK National Facility for Scientific Diving is acknowledged for supporting BAS diving operations. This work was funded by the BAS's core science program, BIOFLAME.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/371/DC1
Materials and Methods
Fig. S1

3 April 2008; accepted 19 May 2008
10.1126/science.1158647

British Antarctic Survey (BAS), Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

*To whom correspondence should be addressed. E-mail: dkab@bas.ac.uk

Identification of SLEEPLESS, a Sleep-Promoting Factor

Kyunghee Koh,[1]* William J. Joiner,[1]* Mark N. Wu,[2]* Zhifeng Yue,[1] Corinne J. Smith,[1] Amita Sehgal[1]†

Sleep is an essential process conserved from flies to humans. The importance of sleep is underscored by its tight homeostatic control. Through a forward genetic screen, we identified a gene, *sleepless*, required for sleep in *Drosophila*. The *sleepless* gene encodes a brain-enriched, glycosylphosphatidylinositol-anchored protein. Loss of SLEEPLESS protein caused an extreme (>80%) reduction in sleep; a moderate reduction in SLEEPLESS had minimal effects on baseline sleep but markedly reduced the amount of recovery sleep after sleep deprivation. Genetic and molecular analyses revealed that *quiver*, a mutation that impairs *Shaker*-dependent potassium current, is an allele of *sleepless*. Consistent with this finding, Shaker protein levels were reduced in *sleepless* mutants. We propose that SLEEPLESS is a signaling molecule that connects sleep drive to lowered membrane excitability.

Insufficient and poor-quality sleep is an increasing problem in industrialized nations. Chronic sleep problems diminish quality of life, reduce workplace productivity, and contribute to fatal accidents (*1*). Although the biological needs fulfilled by sleep are unclear (*2*), they are likely to be important, because sleep is conserved from flies to humans (*3*–*7*) and prolonged sleep deprivation can lead to lethality (*8*–*10*). Identifying mechanisms that control sleep may lead to novel approaches for improving sleep quality.

Sleep is regulated by two main processes: circadian and homeostatic (*11*, *12*). The circadian clock regulates the timing of sleep, whereas the homeostatic mechanism controls the need for sleep. Homeostatic pressure to sleep increases with time spent awake and decreases with time spent asleep. Homeostatic control is thought to influence sleep under normal (baseline) conditions as well as recovery (rebound) sleep after deprivation. However, the molecular mechanisms underlying homeostatic regulation of sleep have remained unclear.

A powerful approach to unraveling a poorly understood biological process is to conduct unbiased genetic screens to identify molecules required for that process. The *Drosophila* model for sleep is well suited for such an approach, which proved invaluable for elucidation of the molecular basis of the circadian clock. Although several *Drosophila* genes have been implicated in sleep regulation [for example, (*13*–*15*)], only one of these, the gene encoding the Shaker (Sh) K^+ channel, was isolated as a result of a genetic screen (*16*). A mutation in this gene causes one of the shortest-sleeping phenotypes known, validating the use of screens and suggesting that control of membrane excitability is a critical requirement for sleep.

Using a large-scale, unbiased genetic screen, we identified a gene, *sleepless* (*sss*), which is required in *Drosophila* for both normal baseline sleep and rebound sleep after deprivation. We find that *sss* encodes a brain-enriched, glycosylphosphatidylinositol (GPI)–anchored membrane protein. We also show that *quiver* (*qvr*), a mutation causing impaired *Sh*-dependent K^+ current (*17*, *18*), is an allele of *sss*, and that Sh protein levels are reduced in *sss* mutant flies. We propose that the SSS protein signals homeostatic sleep drive by enhancing K^+ channel activity and thus reducing neuronal excitability.

Identification of *sss*. To identify genes involved in sleep regulation, we carried out a forward genetic screen for *Drosophila* mutants with reduced daily sleep. We screened ~3500 mutant lines bearing transposon insertions. A histogram summarizing the daily sleep of these lines is shown in Fig. 1A. We selected for further study the mutant line with the lowest amount of daily sleep, which we named *sleepless* (*sss*). To homogenize the genetic background, we outcrossed this strain five times into an isogenic wild-type strain, *iso31*, a line generated specifically for use in behavioral experiments (*19*). Both daytime and nighttime sleep were severely reduced in both male and female *sss* mutants relative to background controls (Fig. 1B). Indeed, a small percentage of *sss* flies (~9% for both males and females) in our assay did not sleep at all—a phenotype never seen in control flies. To our knowledge, *sss* mutants exhibit the most extreme reduction in daily sleep (>85% for males and >80% for females; Fig. 1C) attributable to a single gene mutation.

Despite this extreme reduction in daily sleep, waking activity (defined as activity counts per minute awake) was not significantly elevated in this mutant (Fig. 1D), which suggests that the mutant is not hyperactive when awake (*20*). The marked decrease in sleep amount was largely due to a sharp reduction in the duration of sleep bouts (Fig. 1E). However, decreased sleep in the *sss* mutant was also attributable in part to a significant reduction in the number of daily sleep bouts (Fig. 1F). These phenotypes are recessive in mutant animals, because flies bearing one copy of the *sss* mutation behaved similarly to background controls (Fig. 1, C to F).

SSS is a brain-enriched, GPI-anchored protein. *sss* mutants bear a P-element insertion (*EY04063*, which we refer to as *P1*) in the open reading frame of a gene designated CG33472 by the *Drosophila* Genome Project. The genomic structure of this gene consists of two noncoding exons and five coding exons, the last of which also contains a 3′ untranslated region (3′UTR) predicted to be ~3.9 kb (Fig. 2A). In addition to the original *P1* insertion line, there is a second line, which we call *P2*, bearing a transposon insertion (*f01257*) in the 3′UTR. The SSS protein is predicted to contain a signal peptide, an N-type glycosylation site, and a potential GPI attachment site (Fig. 2, B and C). SSS is well conserved in other insect species, and there is a potential *Caenorhabditis elegans* homolog (F31F6.8 in Wormbase, 46% similarity for amino acids 51 to 133) but no obvious vertebrate homologs. Nonetheless, there may be functional vertebrate homologs with conserved downstream pathways.

To characterize the SSS protein, we used a peptide antigen to generate an antibody (*21*). This antibody recognized two bands on Western blots of wild-type head extracts that were not detectable in sss^{P1} mutant extracts (Fig. 2D), which suggests that sss^{P1} is a severe hypomorph or null allele. Because SSS contains a consensus site for N-type glycosylation, we deglycosylated proteins from head extracts and examined SSS mobility by Western blotting. Under these conditions, only a single band of a lower apparent molecular weight than the two untreated bands was detectable (Fig. 2D), indicating that SSS is glycosylated in vivo.

Because *sss* also contains a potential GPI attachment site, we next examined subcellular localization of SSS. Transfection of *Drosophila* S2R$^+$ cells with a wild-type *sss* construct and staining with the SSS antibody under nonpermeabilizing conditions revealed a subset of the SSS protein expressed on the cell surface (Fig. 2E). Treatment of the cells with phosphatidylinositol-specific phospholipase C (PI-PLC) resulted in severe reduction of surface expression (Fig. 2F) and release of the SSS protein into the culture medium (Fig. 2G). These results show that the SSS protein is attached to the extracellular surface of the plasma membrane with a GPI anchor and can be released by cleavage with PLC.

Using our SSS antibody, we found that SSS protein levels are enriched in fly brain and head relative to body (Fig. 2H). Consistent with these findings, *sss* mRNA expression is enriched by a factor of 23 to 42 in brain relative to whole fly [Adult *Drosophila* Gene Expression Atlas (*22*)]. SSS protein levels did not cycle in a circadian fashion, nor did they change after sleep deprivation (fig. S1, A and B) (see below).

Genetic analysis of *sleepless*. To determine whether the sleep phenotype maps to the *sss* locus, we crossed sss^{P1} to two deficiencies that remove the locus. As predicted, both deficiencies

---

[1]Howard Hughes Medical Institute, Department of Neuroscience, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Division of Sleep Medicine, Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: amita@mail.med.upenn.edu

failed to complement the short-sleeping phenotype of *sss*^{P1} (fig. S2, A and B). To confirm that the sleep phenotype in *sss*^{P1} mutants is caused by disruption of the *sss* gene, we mobilized the P-element to generate precise and imprecise excision lines. Precise excision of the P-element restored daily sleep amount in *sss* mutants to wild-type levels (Fig. 3A and fig. S2C). We also obtained an imprecise excision allele (Δ*40*) that removes part of the *sss* coding region and is likely to be a null allele (fig. S2D). Consistent with this interpretation, *sss*$^{\Delta 40}$ mutants produced an undetectable level of the SSS protein (Fig. 3B). Sleep in this mutant is reduced as severely as in the *P1* mutant; the phenotype maps to the *sss* gene, because the Δ*40* allele failed to complement the *P1* allele (Fig. 3A and fig. S2C).

We next tested whether expression of wild-type SSS from a transgene could rescue the sleep phenotype of *sss*^{P1} mutants. Daily sleep amount was fully rescued to wild-type levels in *sss*^{P1} mutants carrying a genomic *sss* transgene (Fig. 3C and fig. S2E). Together with the results of the deficiency and excision experiments, the rescue data provide strong evidence that disruption of the *sss* gene is responsible for the marked reduction in sleep in *sss*^{P1} mutants.

As described above, *sss*^{P2} mutants harbor an independent transposon insertion in the 3′UTR of the *sss* gene. Homozygous *sss*^{P2} mutant females had amounts of daily sleep similar to those of controls, whereas mutant males had slightly lower amounts of sleep than controls (Fig. 3D and fig. S2F). In contrast, *sss*^{P2}/*sss*^{P1} trans-heterozygous mutants had a ~30% reduction in daily sleep relative to control/*sss*^{P1} flies. These data suggest that the *P2* insertion is a weaker allele than the original *P1* insertion. To examine the biochemical basis of this possibility, we performed Western analysis on head lysates from mutant and control flies. As noted above, the *P1* insertion severely reduced baseline sleep and rendered SSS undetectable (Figs. 2D and 3B). In contrast, the *P2* insertion, which had a minimal effect on baseline sleep, caused a moderate reduction in the level of SSS protein relative to control flies (Fig. 3E). Finally, trans-heterozygous *sss*^{P1}/*sss*^{P2} flies, which exhibit a ~30% reduction in sleep, had a greatly reduced but still detectable level of SSS protein. These data suggest that the amount of daily sleep is correlated with the level of SSS protein and that large reductions of SSS protein are necessary to cause a substantial change in daily sleep.

Reduced homeostatic response in *sss* mutants. We next sought to determine whether *sss* mutants have defects in their homeostatic response to sleep deprivation. We did not observe rebound sleep in *sss*^{P1} flies, but *sss*^{P1} flies do not have much sleep to deprive. Thus, we tested *sss*^{P1}/*sss*^{P2} trans-heterozygous flies, which still have moderate amounts of sleep, as well as *sss*^{P2} homozygotes, which have essentially normal amounts of sleep.

Mechanical stimulation resulted in equivalent sleep loss in *sss*^{P2} homozygous flies and controls; sleep loss was moderately reduced in *sss*^{P1}/*sss*^{P2} flies relative to controls (Fig. 4A and fig. S3A). Whereas control flies showed substantial rebound sleep after deprivation, *sss*^{P1}/*sss*^{P2} flies had little or none (Fig. 4B and fig. S3B). Unexpectedly, we observed a similar lack of rebound sleep in *sss*^{P2} homozygous flies. In addition, when lights were turned on, control animals went to sleep faster after deprivation, but this effect was significantly less pronounced or nonexistent in *sss*^{P2} and *sss*^{P1}/*sss*^{P2} mutants (Fig. 4C and fig. S3C).

Although other genes have been suggested to play a role in homeostatic regulation of sleep, assessment of rebound sleep in animals bearing mutations in these genes is often confounded by concomitant reductions in baseline sleep (*13*, *16*, *23*, *24*). The amount of rebound sleep generally increases with sleep lost (*25*, *26*). Thus, when comparing the effects of sleep deprivation in animals with different amounts of baseline sleep (which leads to loss of different amounts of sleep), it is unclear whether rebound sleep should be compared in absolute terms or relative to amount of sleep lost. We have circumvented this problem by using the *sss*^{P2} mutant to study the contribution of SSS to sleep homeostasis. The finding that *sss*^{P2} animals exhibit markedly reduced rebound sleep, but minimally affected baseline sleep, provides strong evidence that sleep homeostasis is impaired in these mutants.

Effect of *sss* on other behaviors and longevity. To further characterize *sss* mutants, we examined several other behavioral phenotypes. Because mutations in certain central clock genes cause baseline and rebound sleep phenotypes (*9*, *27*–*30*), we analyzed the circadian rhythm phenotypes of *sss* mutants. Whereas *sss*^{P1} mutants exhibited weak rhythms, almost all *sss*^{P1}/*sss*^{P2} trans-heterozygous mutants, which displayed a ~30% reduction in daily sleep time, were rhythmic (Fig. 5, A and B, and table S1). Furthermore, daily oscillations in the level of PERIOD (PER) protein in the ventral lateral neurons (clock cells) remained intact in *sss*^{P1} mutants (Fig. 5C), which suggests that the reduced behavioral rhythmicity seen in these mutants is not due to a defect in the central clock.

Several other behaviors that we tested also appear normal. We found that the phototactic

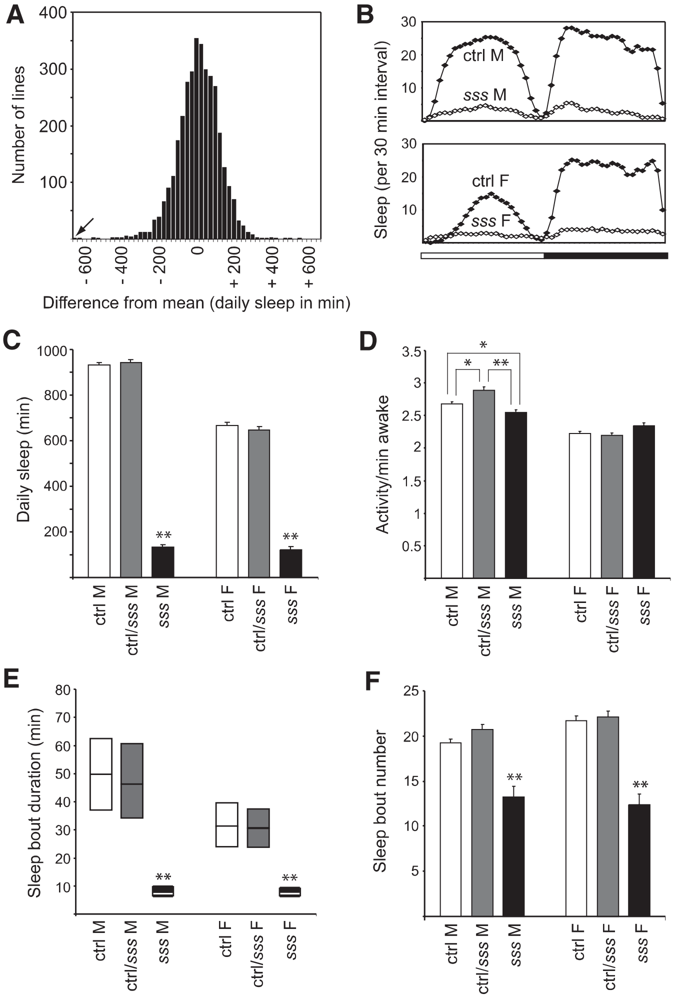


Fig. 1. Sleep phenotype of *sss* mutants. (**A**) Histogram showing the distribution of daily sleep for ~3500 mutant lines (~8 female flies per line). For each line, daily sleep is shown as the difference from the mean of a group of about 100 to 250 lines tested simultaneously. The arrow indicates the *sss* mutant line. (**B**) Sleep profile in 30-min intervals for *sss* flies (open diamonds) versus background controls (ctrl, solid diamonds). Data for male (M) and female (F) flies are shown. The bar below the *x* axis indicates 12-hour light and 12-hour dark periods. (**C**) Daily sleep amount for control (162 males and 148 females), control/*sss* (111 males and 113 females), and *sss* flies (146 males and 148 females). Data from the same flies are shown in (C) to (F). (**D** to **F**) Activity counts per minute awake (D), sleep bout duration (E), and daily number of sleep bouts (F) for male and female control, control/*sss*, and *sss* flies. In this and subsequent figures, error bars represent SEM. *$P < 0.05$, **$P < 0.0001$. For (C), (E), and (F), significance level is shown for *sss* mutants compared to both control and control/*sss* flies. For (D), significance level is shown for pairwise comparisons as indicated by lines. In (E), sleep bout duration, which is not normally distributed, is presented as simplified box plots. The line inside each box indicates the median; the top and bottom represent 75th and 25th percentiles, respectively. About 9% of flies exhibiting zero sleep were excluded from calculation of sleep bout duration.

responses of sss^{P1} mutants are similar to those of controls (fig. S4A) and that sss^{P1} mutants perform as well as controls in a taste discrimination assay (fig. S4B). sss^{P1} flies ($n = 43$) did not exhibit a bang-sensitive paralytic phenotype, whereas 89% ($n = 56$) of *easily shocked* (eas^1) flies used as a positive control did exhibit this phenotype. On the other hand, the sss^{P1} mutants appeared somewhat uncoordinated, and fewer mutants were able to climb a specific distance in given amounts of time relative to controls (fig. S4C). However, despite their apparent difficulties with coordination, sss^{P1} mutants spent more time walking than controls and were capable of flying and mating. Consistent with the widely held view that sleep serves essential biological functions, sss^{P1} mutants also exhibited a shortened life span relative to background controls (Fig. 5D and fig. S5).

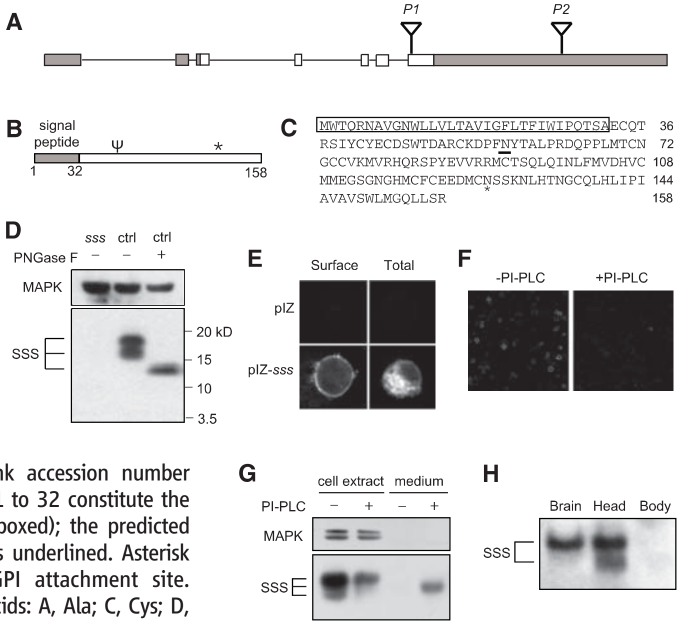


Fig. 2. *sss* encodes a brain-enriched, GPI-anchored protein. (**A**) Schematic of the genomic structure of the *sss* locus. Noncoding regions of the cDNA are shaded; coding regions are shown in white. (**B**) Schematic of structural features of the SSS protein. The primary sequence contains a predicted signal peptide, an N-type glycosylation site (Ψ), and a potential GPI attachment site (*). (**C**) Amino acid sequence of SSS (GenBank accession number EU816195). Amino acids 1 to 32 constitute the predicted signal peptide (boxed); the predicted N-type glycosylation site is underlined. Asterisk denotes the predicted GPI attachment site. Abbreviations for amino acids: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. (**D**) Glycosylation of the SSS protein. Western blot analysis with SSS antibody revealed two bands detected in head extracts from wild-type (ctrl) but not *sss* flies. Deglycosylation of head extracts by treatment with PNGase F resulted in detection of a single band. Because our antibody to SSS does not recognize glycosylated SSS well, Western blots were treated with peptide *N*-glycosidase F (PNGase F) before being probed with antibody to SSS. In this and subsequent Western blots, antibody to mitogen-activated protein kinase (MAPK) was used to control for loading. (**E**) Surface expression of SSS in cultured *Drosophila* cells. $S2R^+$ cells were transfected with a pIZ-*sss* construct and stained with or without permeabilization to assay for total or surface expression, respectively. Transfection with the pIZ vector alone shows specificity of our SSS antibody. (**F**) Reduced surface expression of SSS after PI-PLC treatment. $S2R^+$ cells transfected with a pIZ-*sss* construct were stained without permeabilization after PI-PLC (+) or mock (–) treatment. (**G**) Release of SSS into the culture medium by PI-PLC. Western blot analysis of $S2R^+$ cells transfected with pIZ-*sss* was performed after PI-PLC (+) or mock (–) treatment. (**H**) Enrichment of SSS expression in brain and head versus body. An equal amount of total protein (~40 μg) was loaded per lane. The experiments in (D) through (H) were performed three or four times with similar results.

***sss* is allelic to *qvr* and affects Sh expression.** Because two short-sleeping mutants, *Sh* and *Hyperkinetic* (*Hk*), exhibit ether-induced leg shaking, we assayed *sss* mutants for this phenotype (*16*, *23*). We found that both sss^{P1} and sss^{P2} mutants show ether-induced leg shaking. Notably, *qvr*, a mutant for which the underlying molecular defect is unknown, also has a leg-shaking phenotype, and this phenotype has been mapped close to *sss* (*17*). Because *qvr* mutants exhibit impaired *Sh*-dependent K^+ current (*18*), identification of *qvr* as an allele of *sss* would implicate *Sh* as an effector of SSS function.

Genetic and molecular analyses confirmed that *qvr* is indeed an allele of *sss*. The *qvr* mutation failed to complement sss^{P1} for the leg-shaking phenotype. Similarly, after being outcrossed five times, *qvr* mutants showed a significant decrease in sleep relative to wild-type controls, and sss^{P1}/*qvr* trans-heterozygotes showed a further reduction in sleep (Fig. 6A and fig. S6).

We next investigated the molecular basis of the *qvr* mutation. Reverse transcription polymerase chain reaction (RT-PCR) of *sss* transcripts in *qvr* mutants produced three bands, whereas that of wild-type *sss* transcripts produced a single band (Fig. 6B), indicating splicing defects in *qvr* mutants. None of the three *qvr* bands showed the same electrophoretic mobility as the wild-type control band. Sequencing of the RT-PCR products revealed altered splicing of the last intron (intron 6) of *sss* in the *qvr* mutant (Fig. 6C). A single base change found in the intron is likely to be responsible for the defective splicing (Fig. 6D). Only one of the three *qvr* transcripts (*qvr* 2) is predicted to be in frame (resulting in an insertion of 21 amino acids) and thus has the potential to produce functional SSS protein. Western analysis of *qvr* mutants revealed a small amount of SSS with a slightly higher apparent molecular weight than wild-type SSS protein, which may correspond to the product of the in-frame *qvr* 2 transcript (Fig. 6E).

Because *qvr* mutants were shown to have severely reduced *Sh*-dependent K^+ current (*18*), we examined whether Sh protein levels are affected in *sss* mutants. We found that one form of

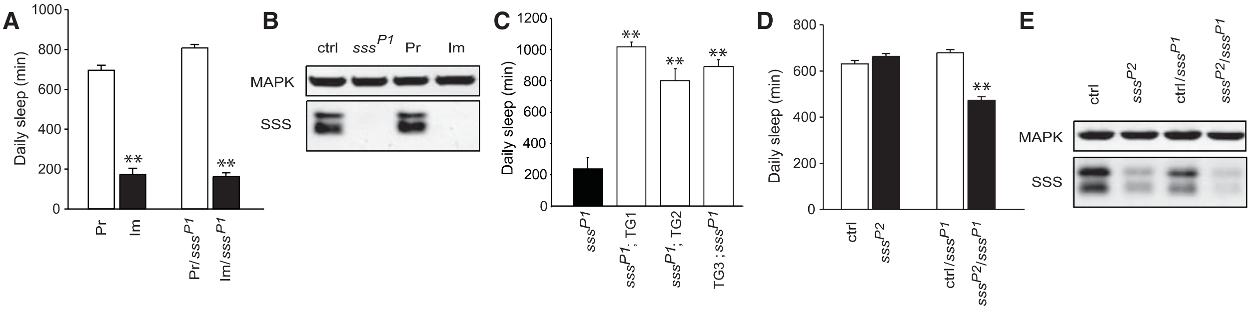


Fig. 3. Genetic analysis of *sss*. (**A**) Daily sleep amount for precise excision (Pr, $n = 26$), $sss^{\Delta 40}$ imprecise excision (Im, $n = 15$), precise/sss^{P1} (Pr/sss^{P1}, $n = 24$), and imprecise/sss^{P1} (Im/sss^{P1}, $n = 35$) female flies. (**B**) Western blot analysis of SSS protein levels. Similar levels of SSS protein are seen in head extracts from background control (ctrl) and precise excision (Pr) flies. SSS protein is undetectable in sss^{P1} and $sss^{\Delta 40}$ imprecise excision (Im) flies. Similar results were obtained in two additional experiments. (**C**) Daily sleep amount for female sss^{P1} mutant flies with (TG1, $n = 15$; TG2, $n = 8$; TG3, $n = 16$) or without ($n = 16$) a genomic *sss* transgene. TG1, 2, and 3 refer to three independent transgene insertions, and one or two copies of the transgene were present in the flies tested. (**D**) Daily sleep amount for sss^{P2} ($n = 110$) versus background control (ctrl, $n = 80$) as well as control/sss^{P1} ($n = 80$) versus sss^{P2}/sss^{P1} ($n = 112$) female flies. (**E**) Reduced levels of SSS protein in sss^{P2} and trans-heterozygous sss^{P2}/sss^{P1} flies. Similar results were obtained in three additional experiments. Data from male flies of the genotypes shown in (A), (C), and (D) are shown in fig. S2. *$P < 0.05$, **$P < 0.0001$.

Sh protein is expressed at a substantially reduced level in sss^{P1} mutants relative to wild-type flies (Fig. 6F), which suggests that SSS affects Sh at least in part through its protein expression. These results establish SSS as an important regulator of the Sh K$^+$ channel.

Discussion. We have identified a *Drosophila* gene required for homeostatic regulation of sleep under normal conditions and after sleep deprivation. Although genes have been identified that regulate sleep-wake stability and baseline sleep amount, few have been shown to be important for sleep rebound (*13*, *15*, *31–35*). Thus, further analysis of SSS function may provide a rare opportunity to gain mechanistic insight into the homeostatic regulation of sleep.

It is worth noting that sss^{P2} animals show a moderate reduction in SSS protein and a minimal reduction in baseline sleep, but have severely reduced sleep rebound. The differential requirement for SSS protein in normal versus rebound sleep may be explained in the context of the two-process model of sleep regulation, where sleep is postulated to be controlled by the opposing influences of circadian waking drive and homeostatic sleep drive (*11*, *36*). In this context, for early-morning rebound sleep to occur, a strong homeostatic signal promoting sleep would be required to counteract a strong circadian input keeping the flies awake. At night, when circadian waking drive is weaker or absent, a relatively low level of homeostatic input may suffice to allow flies to sleep. The moderate level of SSS protein in sss^{P2} mutants may be within the range where sleep is possible when a wake-promoting circadian signal is low (at night), but not when it is high (in the early morning). In contrast, sss^{P1} and $sss^{\Delta 40}$ mutants, which have undetectable levels of SSS expression, display severe reductions in both baseline and rebound sleep. In these mutants, the sleep-promoting signal may be too low to allow flies to sleep even when the circadian waking drive is weak at night.

Clues to the role of SSS at the cellular level come from our biochemical characterization of this molecule. The SSS protein is a GPI-anchored membrane protein enriched in the brain. GPI-anchored proteins can function as ligands or coreceptors and can also act as diffusible signals after cleavage of the GPI anchor (*37*, *38*). Although we were unable to detect circadian or homeostatic regulation of the total levels of SSS protein, such regulation may occur at the level of cleavage of the GPI anchor. Regulation of release is known to be controlled by time of day for other proteins that do not cycle in overall levels, such as pigment-dispersing factor, a molecular output of clock neurons (*39*). Alternatively, SSS may be regulated in a subset of cells that express it, which would be undetectable on our Western blots.

A potential mechanism by which SSS regulates sleep is suggested by our finding that *qvr* is an allele of *sss* and that Sh protein levels are reduced in *sss* mutants. Furthermore, *qvr* mutants exhibit markedly impaired *Sh*-dependent K$^+$ current at the larval neuromuscular junction (*18*). Thus, we propose that SSS lowers membrane excitability by modulating K$^+$ channel expression and activity. It is striking that among thousands of mutants screened in *Drosophila*, two with the strongest sleep phenotypes affect the *Sh* K$^+$ channel (*16*) and its putative regulator, *sss*. Reduced membrane excitability may thus be a central feature of sleep. Collectively, our data suggest that SSS is a signaling molecule that links homeostatic sleep drive to neuronal excitability.

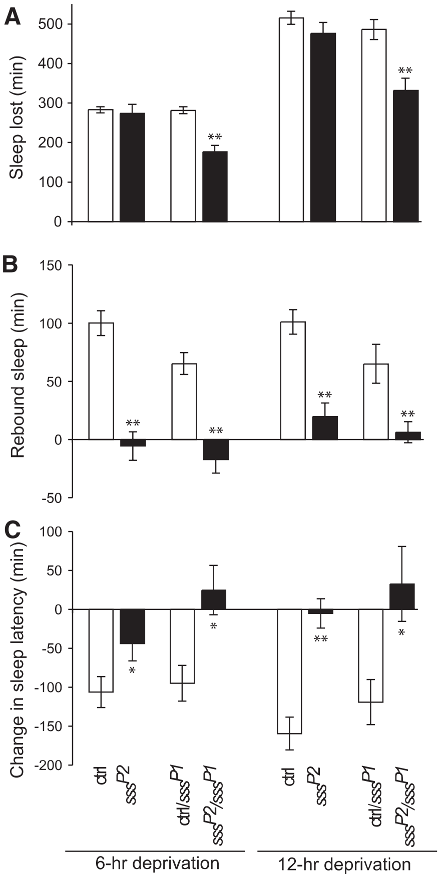


Fig. 4. Reduced homeostatic response to sleep deprivation in female *sss* mutants. (**A**) Amount of sleep lost during 6 or 12 hours of deprivation by the end of the dark period for background control (ctrl), sss^{P2}, control/sss^{P1}, and sss^{P2}/sss^{P1} flies. Data from 13 to 56 female flies are shown. (**B**) Amount of sleep gained during 6 hours of recovery after deprivation as in (A). (**C**) Change in sleep latency after deprivation relative to undisturbed controls as in (A). Sleep latency is defined as the time between the end of deprivation (which coincided with light onset) and the start of a sleep bout. Data from male files are shown in fig. S3. *$P < 0.05$, **$P < 0.001$.

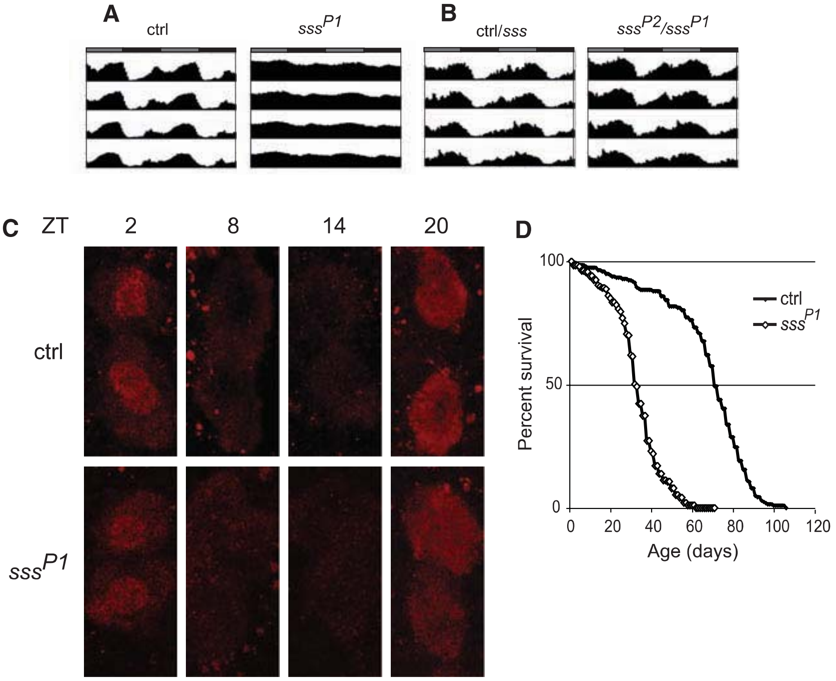


Fig. 5. Circadian rhythm and longevity phenotypes of *sss* mutants. (**A**) Average activity records for background control (ctrl, $n = 64$) and sss^{P1} male flies ($n = 81$) assayed in constant darkness (DD). The activity records are double-plotted so that each horizontal line represents data for 2 days. The gray and black bars above each activity record indicate subjective day and night, respectively. (**B**) Activity records showing average activity in DD for control/sss^{P1} and control/sss^{P2} ($n = 76$) versus sss^{P2}/sss^{P1} ($n = 65$) male flies. Circadian data for control/sss^{P1} and control/sss^{P2} flies were statistically similar and thus were pooled. (**C**) Cycling of PER protein in large ventral lateral neurons in control and sss^{P1} mutants. Ventral lateral neurons for control and sss^{P1} flies were stained for PER at indicated Zeitgeber times (ZT). PER protein levels are elevated at ZT2 and ZT20 and are low at ZT8 and ZT14. (**D**) Survivorship curves of background control (solid diamonds) and sss^{P1} (open diamonds) flies. Female *sss* flies ($n = 187$) show a significantly shorter life span ($P < 0.0001$) than controls ($n = 198$). Data from male flies are shown in fig. S5.

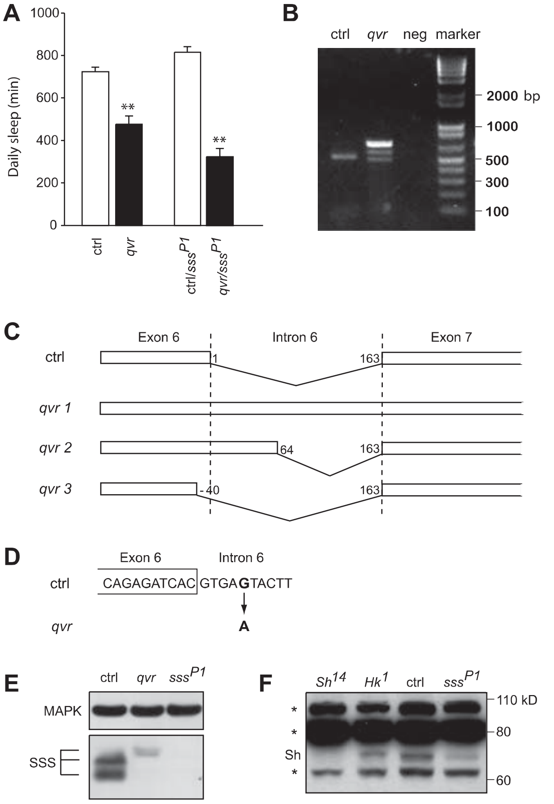


Fig. 6. *sss* is allelic to *qvr* and affects Sh expression. (**A**) Daily sleep amount for *qvr* ($n = 31$), versus background control ($n = 32$) as well as control/sss^{P1} ($n = 30$) versus *qvr*/sss^{P1} ($n = 32$) female flies. **$P < 0.0001$. (**B**) Altered *sss* transcripts in *qvr* mutants. RT-PCR products were obtained with *qvr* and background control (ctrl) RNA and water was used as a negative control (neg). (**C**) Schematic representation of *sss* transcripts in *qvr* mutants. *qvr* 1, 2, and 3 correspond to the top, middle, and bottom bands, respectively. In background control transcripts, 163 nucleotides of intron 6 are spliced out. In contrast, the entire intron is present in *qvr* 1 transcripts. In *qvr* 2 and 3 transcripts, splice donor sites differ from the one used in wild-type control transcripts, as indicated by the nucleotide numbers for splice sites. (**D**) Sequence change in *qvr* genomic DNA in intron 6 of *sss*. The fifth nucleotide in intron 6 has a G → A transition. (**E**) Altered expression of SSS in *qvr* mutants. Fly head extracts from background control, *qvr*, and sss^{P1} flies were analyzed by Western blotting with SSS antibody. (**F**) Reduced expression of Sh in *sss* mutants. Western blot analysis of head extracts with Sh antibody reveals a Sh-specific band that is substantially reduced in sss^{P1} mutants relative to background control flies. Sh^{14} flies were used to identify a Sh-specific band, and Hk^1 flies were used as an additional control. Nonspecific bands (*) may have obscured additional Sh bands. The experiments in (E) and (F) were performed three times with similar results.

References and Notes

1. M. R. Rosekind, *Sleep Med.* **6** (suppl. 1), S21 (2005).
2. J. M. Siegel, *Nature* **437**, 1264 (2005).
3. S. S. Campbell, I. Tobler, *Neurosci. Biobehav. Rev.* **8**, 269 (1984).
4. J. C. Hendricks *et al.*, *Neuron* **25**, 129 (2000).
5. P. J. Shaw, C. Cirelli, R. J. Greenspan, G. Tononi, *Science* **287**, 1834 (2000).
6. R. J. Greenspan, G. Tononi, C. Cirelli, P. J. Shaw, *Trends Neurosci.* **24**, 142 (2001).
7. T. Yokogawa *et al.*, *PLoS Biol.* **5**, e277 (2007).
8. A. Rechtschaffen, M. A. Gilliland, B. M. Bergmann, J. B. Winter, *Science* **221**, 182 (1983).
9. P. J. Shaw, G. Tononi, R. J. Greenspan, D. F. Robinson, *Nature* **417**, 287 (2002).
10. J. L. Pitman, J. J. McGill, K. P. Keegan, R. Allada, *Nature* **441**, 753 (2006).
11. A. A. Borbely, P. Achermann, *J. Biol. Rhythms* **14**, 557 (1999).
12. C. B. Saper, G. Cano, T. E. Scammell, *J. Comp. Neurol.* **493**, 92 (2005).
13. K. Kume, S. Kume, S. K. Park, J. Hirsh, F. R. Jackson, *J. Neurosci.* **25**, 7377 (2005).
14. J. C. Hendricks *et al.*, *Nat. Neurosci.* **4**, 1108 (2001).
15. K. Foltenyi, R. J. Greenspan, J. W. Newport, *Nat. Neurosci.* **10**, 1160 (2007).
16. C. Cirelli *et al.*, *Nature* **434**, 1087 (2005).
17. J. M. Humphreys, B. Duyf, M. L. Joiner, J. P. Phillips, A. J. Hilliker, *Genome* **39**, 749 (1996).
18. J. W. Wang, J. M. Humphreys, J. P. Phillips, A. J. Hilliker, C. F. Wu, *J. Neurosci.* **20**, 5958 (2000).
19. E. Ryder *et al.*, *Genetics* **167**, 797 (2004).
20. R. Andretic, P. J. Shaw, *Methods Enzymol.* **393**, 759 (2005).
21. See supporting material on *Science* Online.
22. V. R. Chintapalli, J. Wang, J. A. Dow, *Nat. Genet.* **39**, 715 (2007).
23. D. Bushey, R. Huber, G. Tononi, C. Cirelli, *J. Neurosci.* **27**, 5384 (2007).
24. W. P. Hu *et al.*, *Sleep* **30**, 247 (2007).
25. R. Huber *et al.*, *Sleep* **27**, 628 (2004).
26. R. Huber, T. Deboer, I. Tobler, *Brain Res.* **857**, 8 (2000).
27. E. Naylor *et al.*, *J. Neurosci.* **20**, 8138 (2000).
28. J. P. Wisor *et al.*, *BMC Neurosci.* **3**, 20 (2002).
29. A. Laposky *et al.*, *Sleep* **28**, 395 (2005).
30. J. C. Hendricks *et al.*, *J. Biol. Rhythms* **18**, 12 (2003).
31. L. Lin *et al.*, *Cell* **98**, 365 (1999).
32. R. M. Chemelli *et al.*, *Cell* **98**, 437 (1999).
33. J. P. Wisor *et al.*, *J. Neurosci.* **21**, 1787 (2001).
34. A. Kramer *et al.*, *Science* **294**, 2511 (2001).
35. D. Kapfhamer *et al.*, *Nat. Genet.* **32**, 290 (2002).
36. D. M. Edgar, W. C. Dement, C. A. Fuller, *J. Neurosci.* **13**, 1065 (1993).
37. M. Hattori, M. Osterfield, J. G. Flanagan, *Science* **289**, 1360 (2000).
38. G. Paratcha *et al.*, *Neuron* **29**, 171 (2001).
39. J. H. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3608 (2000).
40. We thank S. Artavanis-Tsakonas, C.-F. Wu, and C. Cirelli for fly strains, and Y. He, H. Bellen, and the Bloomington Stock Center for sending stocks for the screen. Supported by NIH grant AG017628 (A.S. and K.K.), a University Research Foundation award from the University of Pennsylvania (K.K.), and a Career Award for Medical Scientists from the Burroughs-Wellcome Foundation (M.N.W.). A.S. is an Investigator of the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/372/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References

31 January 2008; accepted 9 June 2008
10.1126/science.1155942

Properties of Gamma-Ray Burst Progenitor Stars

Pawan Kumar,[1]* Ramesh Narayan,[2] Jarrett L. Johnson[1]

We determined some basic properties of stars that produce spectacular gamma-ray bursts at the end of their lives. We assumed that accretion of the outer portion of the stellar core by a central black hole fuels the prompt emission and that fall-back and accretion of the stellar envelope later produce the plateau in the x-ray light curve seen in some bursts. Using x-ray data for three bursts, we estimated the radius of the stellar core to be $\sim(1-3)\times10^{10}$ cm and that of the stellar envelope to be $\sim(1-2)\times10^{11}$ cm. The density profile in the envelope is fairly shallow, with $\rho \sim r^{-2}$ (where ρ is density and r is distance from the center of the explosion). The rotation speeds of the core and envelope are ~0.05 and ~0.2 of the local Keplerian speed, respectively.

Observations of gamma-ray bursts (GRBs) suggest that the activity at the center of these explosions lasts for several hours (*1*, *2*). The most compelling evidence is provided by three bursts (*3*)—GRBs 060413, 060607A, and 070110—that show a sudden decline in their x-ray light curves (LCs) a few hours after the prompt burst (Fig. 1). The flux decline is by a factor of 10 or more and is much too sharp for the radiation to originate in an external forward shock (FS) (*4*); the most likely explanation is continued activity at the center of the explosion, at least until the time of the decline. Additional evidence for continued activity of the central engine is provided by the x-ray flares seen in many GRBs (*5*–*7*) and also by those bursts whose x-ray and optical afterglow LCs are mutually incompatible with a common origin (*8*, *9*). In fact, central engine activity is implicated whenever the observed flux variability time scale, δt, is much smaller than the time

---

[1] Astronomy Department, University of Texas, Austin, TX 78712, USA. [2] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: pk@astro.as.utexas.edu

elapsed since the onset of the explosion. The reason is that, for a relativistic external shock, causality dictates that $\delta t \gtrsim R/2c\Gamma^2$(here, c is the speed of light), and the time it takes for photons to arrive at the observer from a shock front at radius R that is moving with Lorentz factor Γ is also $\sim R/2c\Gamma^2$ (that is, $\delta t/t \sim 1$ for external shocks).

In this paper, we adopt the collapsar model of GRBs (*10*, *11*), in which the inner part of the progenitor star collapses to a rapidly spinning black hole (BH), and the remaining gas from the star accretes onto the BH and produces an ultra-relativistic jet. With a few plausible assumptions, we show here that x-ray observations of the three bursts mentioned above (GRBs 060413, 060607A, and 070110) may be inverted to infer the structure and rotation rate of their progenitor stars. We selected these particular bursts for this work because they have prominent x-ray plateaus that are almost certainly the result of central engine activity. Thus, for these bursts, we have information on the power generation at the center covering an extended period of time, which enables us to determine the core and envelope structure of the progenitor star; only three bursts met this strict requirement of central engine activity–dominated x-ray plateau. Some of the results we find regarding progenitor star properties, especially the core structure, apply to a much larger sample of bursts, as discussed at the end of this paper. In addition, optical emission was detected for two of the three bursts (Fig. 1). The optical LC is consistent with origin in the FS. Moreover, for a reasonable set of parameters, the amount of x-ray flux produced in the FS is found to be $\sim 10^2$ times smaller than the flux observed during the plateau. Thus, there is consistency between a FS origin for the optical LC and central engine activity for the x-ray emission.

The panel on the left in Fig. 2 shows a schematic GRB x-ray LC, which is based on the data shown in Fig. 1. The panel on the right outlines the basic features of our model and shows how the four phases of the LC are connected to corresponding zones in the progenitor star.

We assumed that the star has a core-envelope structure, as is common in stellar models. The bulk of the mass is in the stellar core. Some of the mass in the core collapses directly to form a spinning BH, and the rest accretes on the BH to produce the prompt GRB emission. Surrounding the inner zone is a second zone, which represents the transition region between the core and the envelope. This zone has a steeply falling density profile, and there is a corresponding rapid decline in the x-ray flux. The outermost zone is a relatively low-density stellar envelope. Accretion of this gas produces the plateau in the LC. The drop in the x-ray flux at the end of the plateau, which can be sudden, corresponds to the time when the outermost layers of the envelope are accreted.

The fall-back time is the time it takes a parcel of gas in the progenitor star at radius r to fall to the center, and it is approximately equal to the free-fall time (*12*), $t_{\rm fb} \sim 2(r^3/GM_{\rm r})^{1/2}$, where G is the gravitational constant, $M_{\rm r}$ is the mass interior to r, and all times are in the frame of the GRB host galaxy. Let us assume that the accretion time $t_{\rm acc}$ of the gas in the disk is smaller than $t_{\rm fb}$ (see below). Then, at a given time $t = 10^2 t_2$ s, the central engine will accrete gas that has fallen back from a radius $r = 10^{10} r_{10}$ cm, where

$$r_{10} \sim 1.5 t_2^{2/3} M_{\rm BH,1}^{1/3} \tag{1}$$

and $M_{\rm BH,1}$ is the BH mass in units of $10M_\odot$, where $M_\odot$ is the mass of the sun. We are assuming here that the interior mass $M_{\rm r}$ is dominated by the BH. Let us suppose that the gas at radius r in the progenitor star has an angular velocity Ω equal to a fraction f_Ω of $\Omega_{\rm k}$, the local Keplerian angular velocity

$$\Omega(r) = f_\Omega(r)\Omega_{\rm k}(r) = f_\Omega(r)(GM_{\rm r}/r^3)^{1/2} \tag{2}$$

Then, the specific angular momentum $j_* = 10^{18} j_{*,18}$ cm^2 s^{-1} of this gas is given by

$$j_{*,18}(r) \approx 3.8 M_{\rm BH,1}^{1/2} r_{10}^{1/2} f_\Omega(r) \approx 4.7 t_2^{1/3} M_{\rm BH,1}^{2/3} f_\Omega(r) \tag{3}$$

The gas will fall back toward the black hole until its angular momentum allows it to become cen-

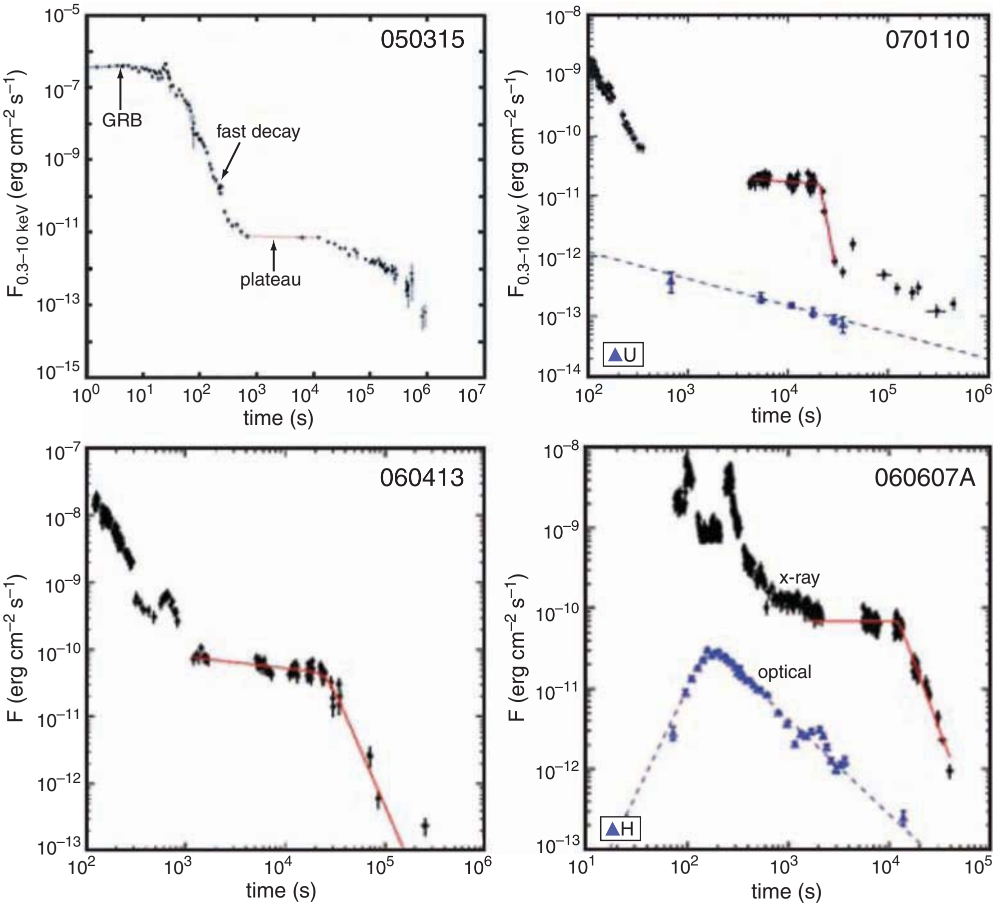


Fig. 1. (**Top left**) 0.3-to-10–keV band x-ray LC of a typical long-duration burst, GRB 050315 [panel is from (*21*)]. Four distinct phases are seen in the LC: (i) the prompt GRB lasting for about a minute, (ii) an early steep decline lasting for ~10 min, (iii) a plateau lasting until ~10^4 s, and (iv) a "normal" post-plateau decay. (**Top right**) X-ray and optical LCs for GRB 070110 [panel is from (*3*)]. The x-ray LC falls very sharply at the end of the plateau ($t \sim 3 \times 10^3$ s) before switching to a normal decline, whereas the optical LC is a single power-law for the entire duration. (**Bottom panels**) X-ray LCs of GRBs 060413 and 060607A [panels are also from (*3*)]. Note the complex x-ray LCs of these bursts, with flares, breaks, and plateaus. In contrast, the optical LCs of GRBs are typically smooth and simple, such as those of GRBs 070110 and 060607A, shown in the right panels in blue.

trifugally supported, at which point it becomes part of a thick accretion disk about the black hole. Its further fall will then occur on the time scale at which viscous forces in the disk allow the orbit to decay.

The viscous accretion time of the fall-back gas after it has circularized is approximately equal to $t_{\mathrm{acc}} \sim 2/\alpha\Omega_{\mathrm{k}}(r_{\mathrm{d}})$, where $\alpha = 0.1\alpha_{-1}$ is the viscosity parameter, and $r_{\mathrm{d}} \sim j_{*}^{2}/GM_{\mathrm{BH}}$ is the radius at which gas with specific angular momentum j_{*} circularizes. Here we have assumed that the accretion disk is geometrically thick, which is generally true at the high mass accretion rates expected in a collapsar (*13*). We thus have

$$t_{\mathrm{acc}} \sim 10\alpha_{-1}^{-1} f_{\Omega}^{3}(r) t_{\mathrm{fb}} \sim 10\alpha_{-1}^{-1} j_{\mathrm{d},18}^{3} M_{\mathrm{BH},1}^{-2}\ \mathrm{s} \qquad (4)$$

The first relation shows that the accretion time will be shorter than the fall-back time, as long as $f_{\Omega} \lesssim 0.4$ (for $\alpha \sim 0.1$).

The jet power from the central engine is determined by the mass accretion rate $\dot{M}_{\mathrm{BH}}$ onto the BH, which is usually a small fraction of the mass fall-back rate $\dot{M}_{\mathrm{fb}}$ onto the accretion disk [the majority of the fall-back mass is expelled in a disk wind as it spirals in toward the BH (*14*)].

If the mass fall-back rate decreases suddenly at some time t_0—e.g., because of the transition from the stellar core to the envelope or when gas at the outer edge of the envelope has fallen back—then it can be shown by the conservation of mass and angular momentum that the accretion rate on the BH will decline with time as (*15*)

$$\dot{M}_{\mathrm{BH}}(t) \sim \dot{M}_{\mathrm{BH}}(t_0)[1 + 1.5(t - t_0)/t_{\mathrm{acc}}]^{-2}, t \geq t_0 \qquad (5)$$

The exponent outside the square brackets is model-dependent and varies between −4/3 and −8/3, depending on details; for definiteness, we choose −2. If $t_{\mathrm{acc}} << t_0$, the effect of the above time dependence is that, in a log-log plot, the jet power will initially drop by a large factor $\sim(t_0/t_{\mathrm{acc}})^2$ within a time $\sim t_0$, and the power will then transition to an asymptotic decline proportional to t^{-2}. This feature in the predicted LC plays an important role in our model. It is how we explain the sudden decline in the x-ray LC at the end of the prompt GRB and at the end of the plateau.

We apply these scaling relations to derive constraints on the stellar structure and rotation rate of GRB progenitors. A number of the features seen in GRBs can be explained naturally within the collapsar model as corresponding to the accretion of different portions of the progenitor star onto the central BH. Here we quantify the size, density, and rotation rate of the stellar core and envelope, as inferred from the observed time scales discussed above.

Observations with the x-ray telescope aboard the Swift satellite show that the flux in the 0.3-to-10–keV band of a typical long-duration GRB is roughly constant for ~20 s (Fig. 1, top left panel), after which the x-ray LC undergoes a rapid decline as $\sim t^{-3}$ or faster for ~5 min. Assuming that it takes ~30 s for the BH to form and adding to this the burst duration of 20 s, we estimated that the end of the main burst is ~50 s after the initiation of core collapse. By Eq. 1, the gas that falls back at this time comes from a radius $\sim 9 \times 10^{9} M_{\mathrm{BH},1}^{1/3}$ cm. This radius must correspond to the outer edge of the innermost zone in the schematic stellar model shown in Fig. 2.

The subsequent steeply declining phase after the initial burst goes from $t \sim 50$ to 300 s. In our model, this portion of the LC is associated with accretion of gas from the second zone in the star, which is the transition region between the core and the envelope where the density has a steep decline. Using the same argument as before, the outer radius of the second zone must be $\sim 3 \times 10^{10} M_{\mathrm{BH},1}^{1/3}$ cm.

A rapid decline in the x-ray LC requires a rapid decline in $\dot{M}_{\mathrm{BH}}$ (faster than $\sim t^{-2}$), which is possible only when the disk is able to adjust quickly to a decrease in $\dot{M}_{\mathrm{fb}}$. For an accretion flow to respond quickly to rapid changes in $\dot{M}_{\mathrm{fb}}$ at time t, we must have $t_{\mathrm{acc}} << t$ (see Eq. 5). This condition provides an upper limit on the rotation speed in the core region of the progenitor star. The x-ray flux typically drops by a factor of $\sim 10^{3}$ during the rapid decline phase at the end of the prompt burst (Fig. 1). Therefore, by Eq. 5 we require $(250/t_{\mathrm{acc}})^{2} \gtrsim 10^{3}$ or $t_{\mathrm{acc}} \lesssim 8\mathrm{s}$ at $t \sim 50$ s. From Eq. 4, this means we must have $f_{\Omega} \lesssim 0.2\alpha_{-1}^{1/3}$ at $r \approx 10^{10}$ cm.

We can obtain a lower limit on f_{Ω} in the stellar core by requiring that the angular momentum should be sufficiently large that the fall-back gas is able to form an accretion disk; i.e., $r_{\mathrm{d}} \gtrsim 3R_{\mathrm{g}}$, where $R_{\mathrm{g}} \equiv GM_{\mathrm{BH}}/c^{2}$ (the radius of the innermost stable circular orbit is $2.3R_{\mathrm{g}}$ for a BH with a spin parameter of 0.9). Because during core collapse a particle initially at r ends up at a radius $r_{\mathrm{d}} \approx r[f_{\Omega}(r)]^{2}$, where it is centrifugally supported, the condition $r_{\mathrm{d}} \gtrsim 3R_{\mathrm{g}}$ yields $f_{\Omega}(r) \gtrsim 0.02$. An additional constraint on the core rotation rate is provided by the total energy of $\sim 10^{52}$ erg observed in a typical GRB. Numerical simulations show that the efficiency for jet production from a rapidly rotating BH is ~1% (*16*), and therefore, the total mass accreted by the BH should be $\sim 0.5M_{\odot}$. This suggests that much of the mass within $r \approx 10^{10}$ cm collapses directly to the BH, and thus, $f_{\Omega}(r)$ is not much larger than 0.02; a reasonable estimate is $f_{\Omega} \sim 0.05$ for $r \lesssim 10^{10}$ cm.

A plateau in the x-ray LC is seen typically from $\sim 3 \times 10^{2} - 3 \times 10^{3}$ s (time in host galaxy rest frame). Gas falling on the central accretion disk during this time interval arrives from $3 \times 10^{10} M_{\mathrm{BH},1}^{1/3}\mathrm{cm} \lesssim r \lesssim 1.5 \times 10^{11} M_{\mathrm{BH},1}^{1/3}\mathrm{cm}$ (Eq. 1). We suggest that the gas accreted during this phase comes from the envelope of the progenitor star, which means that the outer radius of the envelope must be $\sim 1.5 \times 10^{11}$ cm. In principle, the plateau could be produced by supernova ejecta that failed to escape. However, such a model is unlikely to give the sharp cutoff seen in the x-ray LC at the end of the plateau in the three GRBs considered in this paper. For these GRBs at least, the plateau must arise from fall-back of the stellar envelope, which has a well-defined outer edge.

The density profile in the stellar envelope can be determined from the rate of decline of the x-ray flux during the plateau: $f_{\mathrm{x}} \propto t^{-\delta}$, $\delta \sim 0.5$. If the specific angular momentum of the gas in the envelope is constant, as might be the case if the envelope is convective, then the required density profile is $\rho(r) \propto r^{-3(\delta+1)/2} \sim r^{-2.2}$ for $\delta \sim 0.5$ (*15*). For a convective stellar envelope that is not very massive, we expect $\rho \propto r^{-3/2}$ when gas pressure dominates and $\rho \propto r^{-3}$ when radiation pressure dominates. The density profile we infer, assuming a constant specific angular momentum envelope, lies between these two limits. If the specific angular momentum increases with increasing radius, then a shallower decline of density is indicated; for instance, $\rho \propto r^{-1.5}$ if the specific angular momentum increases with distance as $\sim r^{3/4}$. A solid-body rotation profile for the envelope is ruled out because it requires an unphysical density structure in which ρ is almost independent of r.

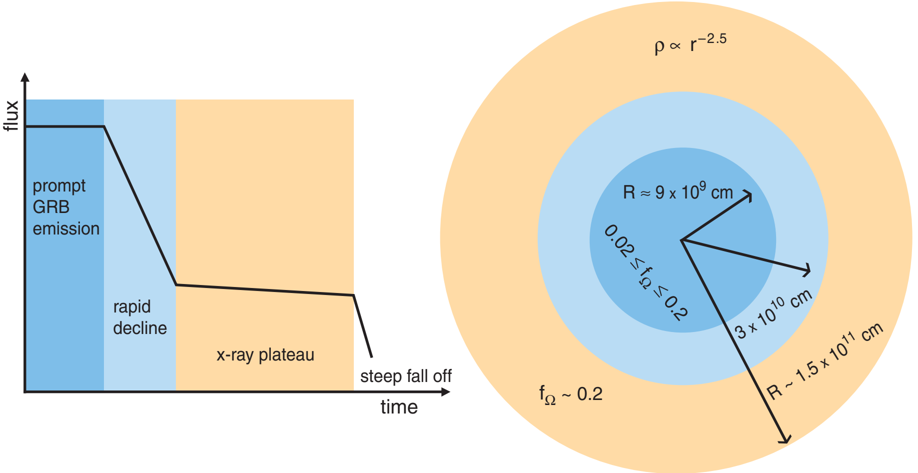


Fig. 2. (**Left**) Schematic x-ray LC with the following four segments: (i) a prompt emission phase, (ii) a steep decline phase, (iii) a plateau phase, and (iv) a post-plateau phase. For the three GRBs considered in this work, the last phase has a steep and sudden decline. (**Right**) Our proposal for how the different segments in the LC are related to the accretion of corresponding zones in the progenitor star. The radii (r) and spin parameters ($f_{\Omega} \equiv \Omega/\Omega_{\mathrm{k}}$) of the various zones were estimated from the x-ray data.

The x-ray flux at the end of the plateau in GRBs 060413, 060607A, and 070110 falls off sharply by a factor ~10^2. The fall-off occurs at $t \sim 3 \times 10^3$ s and extends over a time $\delta t \sim t$. This steep fall-off requires the accretion time to be sufficiently short, $t_{\rm acc} \lesssim t/10 \sim 300$s, which implies that $j_{{\rm d},18} \lesssim 3\alpha_{-1}^{1/3} M_{{\rm BH},1}^{2/3}$ and $f_\Omega(r) \lesssim 0.2\alpha_{-1}^{1/3}$ at $r \sim 1.5 \times 10^{11}$ (*17*, *18*).

A lower limit for f_Ω in the stellar envelope is obtained by the requirement that the in-falling gas should be able to form an accretion disk at the center; this implies that $f_\Omega \gtrsim 0.01$. Another constraint comes from the fact that, apart from the flares discussed below, the x-ray LC during the plateau is usually quite smooth. This suggests that $t_{\rm acc}$ is fairly large (which plays the role of a smoothing time scale), and so f_Ω is probably not much less than the upper limit of ~0.2.

After the sharp fall-off at the end of the plateau, the LC is expected to decline as $\sim t^{-2}$ if the x-ray emission is dominated by jet luminosity associated with the debris disk (see Eq. 5). However, at these low flux levels the observed x-rays might be dominated by emission from shock-heated circumburst gas (i.e., external shock emission). In this case the flux will decline as $\sim t^{-1}$; e.g., GRB 070110 (Fig. 1).

For those GRBs where the x-ray LC makes a smooth transition at the end of the plateau phase, f_Ω should be such that $t_{\rm acc} \sim t$. For these systems, we require that $f_\Omega \sim 0.4$ near the outer edge of the progenitor star.

Rapid flares with short rise times are often seen during the plateau phase of the x-ray LC, and these provide additional constraints. A flare was seen at the beginning of the x-ray plateau in GRBs 060413 and 060607A and at the end of the plateau in GRB 070110 (Fig. 1), with a rise time on the order of $0.1t$. Assuming the flares are produced by a disk instability, the rise time should be a factor of a few larger than the instability time scale $t_{\rm inst}$. For a viscous instability, $t_{\rm inst} \sim t_{\rm acc}$, whereas for a dynamical (e.g., gravitational) instability, $t_{\rm inst} \sim \Omega_{\rm k}^{-1} \sim \alpha t_{\rm acc}$. Taking the flare rise time to be ~5 $t_{\rm inst} \sim 0.1t$, we find that $f_\Omega \sim 0.1\alpha_{-1}^{1/3}$ in the stellar envelope if flares arise as a result of viscous instability, and $f_\Omega \sim 0.3$ for a dynamical instability origin for flares. The amplitude of the flare $f_{\rm x,flare}/f_{\rm x,plateau}$ can be at most $\sim t_{\rm acc}/t_{\rm inst}$. Thus, even in the limit of a dynamical instability, the amplitude is limited to $\sim\alpha^{-1} \sim 10$. Much larger flare amplitudes [for example, in GRB 060526 (*5*)] might suggest an unusually small value of the viscosity parameter α. Alternatively, these flares may be caused by a sudden increase in the mass fall-back rate, though such an event is not easy to visualize in our model.

The decay time of flares is expected to be on the order of $t_{\rm acc}$, as this is the time scale on which a transient enhancement of the accretion rate, regardless of its origin, will subside. A noticeable difference between the x-ray plateau and the prompt emission is that LCs are typically more variable during the burst, and this raises a question as to why the central engine behaves differently during the two phases. Part of the difference is that instability time scales are longer when the outer envelope of the star, with a larger specific angular momentum, is accreted. Another factor is that, during the plateau phase, the jet propagates through an already evacuated cavity and is less prone to fluctuating baryon loading.

An upper limit on the stellar radius R_* is provided by the requirement that the energy expended as the relativistic jet makes its way out of the star not exceed the energy initially injected into the jet. Using results from (*19*, *20*), we find that $R_* \lesssim 5 \times 10^{11}$cm. Our radius estimates are consistent with this limit.

A possible way to get a handle on the mass of the GRB progenitor star is via the total energy produced in the explosion. The mass accreted during the prompt burst was estimated above to be $\sim 0.5 M_\odot$. The energy release during the plateau is typically ~10% of that in the prompt burst; that is, $\sim 10^{51}$ erg (assuming the same beaming factor and efficiency as during the initial burst). Therefore, the mass accreted by the BH is only $\sim 0.03 M_\odot$. Even after allowing for the fact that only a small fraction $\sim (r_{\rm d}/R_{\rm g})^{1/2} \sim 10^{-2}$ of the total fall-back mass actually reaches the BH, the rest being carried away in a disk wind, we estimate that the mass of fall-back gas in the plateau phase is no more than a few $M_\odot$. The mass of the BH can be constrained because much of the mass within $\sim 10^{10}$ cm collapses to the BH; for an evolved star this mass is $\sim 5 M_\odot$. Thus, we estimate the mass of the GRB progenitor star to be $M_* \sim 10 M_\odot$, plus whatever mass is ejected in the accompanying supernova explosion, which we are unable to constrain.

Figure 2 summarizes our primary results on the properties of GRB progenitor stars. The three bursts we considered in this paper are notable in that they have a steep fall-off of x-ray flux at the end of the plateau phase. These bursts provide the most detailed information on the properties of their progenitor stars. Many of our arguments and results, especially those in which we use the prompt burst and subsequent steep fall-off to infer the properties of the stellar core, should apply to any long-duration GRB. Similarly, our discussion of the x-ray plateau can be applied to other bursts that have plateaus in their x-ray LCs that are produced by continuous accretion activity; i.e., those GRBs that show a simple power-law decline of the optical LC and a plateau in the x-ray. These bursts make up roughly one-third of the long GRBs observed by Swift.

References and Notes

1. The possibility of a long-lived central engine activity was suggested (*2*) soon after the discovery of GRB afterglow emission.
2. J. I. Katz, T. Piran, R. Sari, *Phys. Rev. Lett.* **80**, 1580 (1998).
3. E.-W. Liang, B.-B. Zhang, B. Zhang, preprint available at http://arXiv.org/abs/0705.1373 (2007).
4. E. Nakar, J. Granot, preprint available at http://arXiv.org/abs/astro-ph/0606011 (2006).
5. D. N. Burrows *et al.*, *Science* **309**, 1833 (2005); published online 18 August 2005, 10.1126/science.1116168.
6. G. Chincarini *et al.*, preprint available at http://arXiv.org/abs/astro-ph/0702371 (2007).
7. B. Zhang, *Chin. J. Astron. Astrophys.* **7**, 1 (2007).
8. Y. Fan, T. Piran, *Mon. Not. R. Astron. Soc.* **369**, 197 (2006).
9. A. Panaitescu *et al.*, *Mon. Not. R. Astron. Soc.* **369**, 2059 (2006).
10. S. E. Woosley, *Astrophys. J.* **405**, 273 (1993).
11. B. Paczynski, *Astrophys. J. Lett.* **494**, L45 (1998).
12. The factor of 2 in the expression for $t_{\rm fb}$ accounts for the time it takes signals to travel to r and communicate the loss of pressure support at the center.
13. R. Narayan, T. Piran, P. Kumar, *Astrophys. J.* **557**, 949 (2001).
14. K. Kohri, R. Narayan, T. Piran, *Astrophys. J.* **629**, 341 (2005).
15. P. Kumar, R. Narayan, J. L. Johnson, preprint available at http://arXiv.org/abs/0807.0441 (2008).
16. J. C. McKinney, *Astrophys. J. Lett.* **630**, L5 (2005).
17. The decline of the LC cannot be steeper than $\sim t^{-3}$; this is due to late arriving photons from outside the relativistic beaming angle of Γ^{-1} (*18*). Therefore, we can only place an upper limit on f_Ω from the steepness of the decline after the x-ray plateau.
18. P. Kumar, A. Panaitescu, *Astrophys. J. Lett.* **541**, L51 (2000).
19. E. Ramirez-Ruiz, A. Celotti, M. J. Rees, *Mon. Not. R. Astron. Soc.* **337**, 1349 (2002).
20. C. D. Matzner, *Mon. Not. R. Astron. Soc.* **345**, 575 (2003).
21. S. Vaughan *et al.*, *Astrophys. J.* **638**, 920 (2006).

11 April 2008; accepted 17 June 2008
Published online 26 June 2008;
10.1126/science.1159003
Include this information when citing this paper.

High-Resolution Scanning X-ray Diffraction Microscopy

Pierre Thibault,[1]* Martin Dierolf,[1] Andreas Menzel,[1] Oliver Bunk,[1] Christian David,[1] Franz Pfeiffer[1,2]

Coherent diffractive imaging (CDI) and scanning transmission x-ray microscopy (STXM) are two popular microscopy techniques that have evolved quite independently. CDI promises to reach resolutions below 10 nanometers, but the reconstruction procedures put stringent requirements on data quality and sample preparation. In contrast, STXM features straightforward data analysis, but its resolution is limited by the spot size on the specimen. We demonstrate a ptychographic imaging method that bridges the gap between CDI and STXM by measuring complete diffraction patterns at each point of a STXM scan. The high penetration power of x-rays in combination with the high spatial resolution will allow investigation of a wide range of complex mesoscopic life and material science specimens, such as embedded semiconductor devices or cellular networks.

Imaging techniques that rely on coherence, loosely designated as coherent diffractive imaging (CDI) techniques, have thrived in situations where more traditional modes of microscopy are difficult to implement. CDI allows high-resolution imaging while eliminating the need to use high-quality lenses, especially hard to produce in the case of x-rays. This experimental convenience comes, however, at the price of harder data analysis, which often entails reconstructing the image by solving the so-called phase problem. The simplest CDI method, called diffraction microscopy, involves the reconstruction of an image or a three-dimensional (3D) density from far-field diffraction intensities. In the past decade, a number of successful demonstrations of this technique have been reported, ranging from fabricated nanostructures in 2D (*1*, *2*) and 3D (*3*, *4*) to biological specimens in 2D (*5*, *6*). Requirements such as the complete isolation of the specimen have promoted the development of alternative approaches, such as Fourier holography (*7*, *8*) and "ptychography" (*9*–*13*). The latter is an experimental method developed in the 1970s for electron microscopy (*14*), which consists of measuring multiple diffraction patterns by scanning a finite illumination on an extended specimen. Overlap between adjacent illumination positions provides overdetermination in the data. Assuming the profile of the illumination is known, solving the phase problem has been demonstrated to be easier.

Another powerful microscopy technique is scanning transmission x-ray microscopy (STXM). It consists in scanning a focused x-ray beam on a specimen and measuring the transmitted intensity at each raster point. It thus provides a map of the specimen's transmission. Although more advanced imaging modes exist (*15*, *16*), common to all of them is that the resolution is limited by the spot size on the specimen. Increase in resolution thus largely depends on improving focusing optics. STXM relies partly on coherence—higher coherence allows smaller spot sizes—but has developed quite independently from CDI methods.

In this work, we have designed a technique that we call scanning x-ray diffraction microscopy (SXDM), which bridges the gap between scanning microscopy and coherent diffractive methods. To demonstrate this method experimentally, we imaged a buried nanostructure by using a focused hard x-ray beam. The apparatus, described below, is similar to the experimental setups used in both fields. Data acquisition was greatly facilitated by the use of Pilatus, a fast single-photon counting detector having no read-out noise (*17*). In essence, the experiment follows closely the configuration in which ptychography was first developed. We used a separate reconstruction method (*18*) to complement the standard STXM analysis tools and reached resolutions about five times higher than the focal spot dimensions. Our algorithmic approach overcomes the main difficulty of ptychographic measurements by automatically recovering the incident illumination profile at the same time as the complex-valued image of the specimen. This feature is critical to obtain high-quality reconstructions because the detailed structure of the wavefield incident on the specimen is difficult to extract by other means.

A schematic view of the experimental setup is shown in Fig. 1A. An incoming coherent x-ray beam is focused down to a few hundred nanometers on a specimen. The intense focal spot, called the probe, interacts with a small portion of the specimen before propagating out to a pixel array detector in the far-field region. One complete data set is produced by recording diffraction patterns as the specimen is scanned through the focal spot, typically on a 2D grid.

Each measured diffraction pattern carries information on the specimen. More precisely, the wave at the exit of the specimen corresponding to the jth diffraction data set can be expressed as

$$\psi_j(\mathbf{r}) = P(\mathbf{r}-\mathbf{r}_j)O(\mathbf{r}) \qquad (1)$$

where $P(\mathbf{r}-\mathbf{r}_j)$ is the probe wavefield translated by a known amount $\mathbf{r}_j$, and $O(\mathbf{r})$ is the specimen's transmission function, called the object (*19*). Exit waves ψ_j are called views because they effectively provide information on selected regions of the specimen. Redundancy in the data, needed for ptychographic applications, is enforced with a scanning step size small enough for adjacent views to overlap.

To demonstrate the technique, we measured a SXDM data set, using as a specimen a Fresnel zone plate buried under a gold layer (*20*). Imaging this type of nanofabricated device demonstrates the penetration power of x-rays and allows a reliable assessment of the resolution and the validity of a reconstruction. The complete data set is a raster scan of 201-by-201 diffraction patterns, produced with a 6.8-keV x-ray beam focused to a 300-nm spot size, each exposed for 50 ms (*18*). Four of these diffraction patterns are shown in Fig. 1B. These diffraction patterns were sampled on a region of the detector that subtended an angular range of 10 mrad by 10 mrad.

As immediate feedback to the microscopist to define regions of interest, for instance, the SXDM data set can be analyzed by using standard scanning microscopy techniques (*15*, *16*). A simple calculation of the zeroth and first moments of the intensity distribution in each diffraction pattern gives low-resolution images of the transmission and of the gradient of the phase shift produced as the wave traverses the specimen (*21*).

The result of this STXM analysis on a SXDM data set is shown in Fig. 2, B to D. Whereas the spot size of 300 nm limits the resolution of these preliminary micrographs, it takes on quite a different role in the following analysis, namely, to ensure that the diffraction pattern is sufficiently "oversampled" by the detector in the far field.

The approach we have adopted to reconstruct a high-resolution image from a SXDM data set departs from both earlier methods, that is, the Wigner deconvolution (*9*, *11*) and the ptychographic iterative engine (*22*, *23*). We used the difference map (*24*), an iterative algorithm now known to have a wide range of applications going far beyond imaging (*25*). Problems solved by the difference map are formulated as a search for the intersection between two constraint sets. Any solution attempt is represented by a state vector, which typically satisfies neither constraint until the solution is found. For phase retrieval, the state vector generally is a 2D image or a 3D density. For SXDM reconstructions, it is given by the collection of views corresponding to each position of the probe relative to the object. The two constraints on the state vector are the Fourier and the overlap constraints. The former enforces consistency of the views with their corresponding diffraction patterns (*26*). The latter is given by Eq. 1 and formalizes the redundancy of information in the data set, which gives this algorithm, like related ptychographic techniques (*9*, *11*, *22*, *23*), its robustness. Projections onto both constraint sets can be precisely defined and easily implemented in a computer program.

At each iteration, the current state is updated through a combination of the two projections, as prescribed by the difference map algorithm. Each computation of the overlap projection yields new estimates of both the probe and the object (*27*). Unlike many diffraction microscopy reconstructions, this type of problem typically features very short convergence times, a few tens of iterations. It should also be noted that SXDM reconstructions do not suffer from the defocus ambiguity that is met in other CDI methods. The multiplicative relation between the probe and the object ensures that the plane of the reconstruction is well defined.

A reconstruction example, from a 61-by-61 subset of the 201-by-201 diffraction patterns data set, is given in Fig. 3. A second region of interest, closer to the center of the zone plate, is shown in fig. S1. The retrieved complex-valued image provides a map of the optical transfer function of the

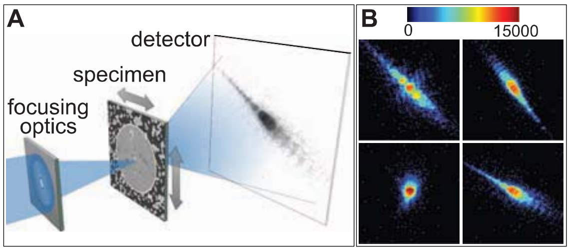


Fig. 1. Schematic representation of the experimental setup. (**A**) The incoming hard x-ray beam is focused on the specimen. A pixel array detector records a complete oversampled diffraction pattern for each raster position of the specimen. (**B**) Examples of diffraction patterns. The color scale is logarithmic, and the units are in photon counts. The focused beam geometry generates a diverging beam past the specimen, thereby removing the need for a beam stop at the detector.

---

[1]Paul Scherrer Institut, 5232 Villigen PSI, Switzerland. [2]École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.

*To whom correspondence should be addressed. E-mail: pierre.thibault@psi.ch

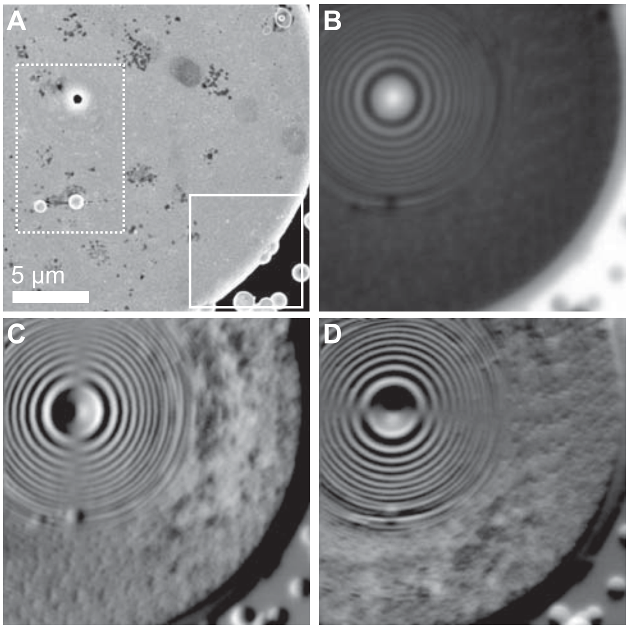


Fig. 2. Preliminary analysis of a measured data set of 201-by-201 diffraction patterns. This test specimen is a Fresnel zone plate of 30-µm diameter and with 70-nm outer zone width. (**A**) SEM image. The zone plate structure is completely buried under a layer of gold. A full reconstruction of the region of interest (solid line) is shown in Fig. 3. The reconstructed image of the dashed line region is shown in fig S1. (**B**) Transmission of the specimen, obtained from the total transmitted intensity at each probe position. (**C**) Phase gradient of the transmitted wave in the horizontal direction. (**D**) Phase gradient in the vertical direction. (B) to (D) involve no phase retrieval and can be produced as data are acquired. The pixel size is equal to the scanning grid spacing (100 nm), but the resolution is limited by the size of the probe (about 300 nm).

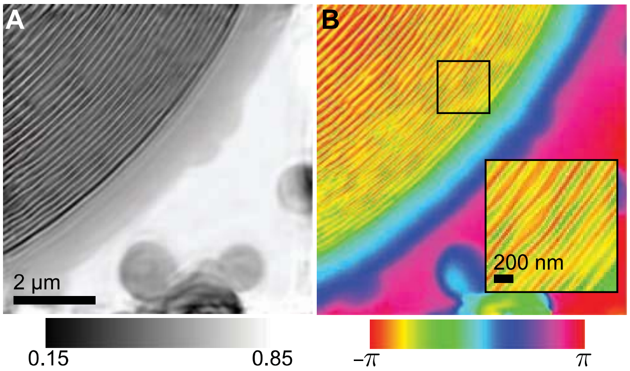


Fig. 3. SXDM reconstruction of the complex optical transmission function of the zone plate specimen. (**A**) Amplitude and (**B**) phase of the reconstruction of a selected region of 61-by-61 diffraction patterns, from the same data set that was used for Fig. 2. The magnified region shows imperfections in the nanofabricated zones. Other defects, corroborated by a SEM image, are shown in fig. S1. The outermost rings of the zone plate (70 nm wide) are very well resolved. The pixel size is 18 nm. The total acquisition time for this image is 186 s, with a total dose of 2×10^9 photons.

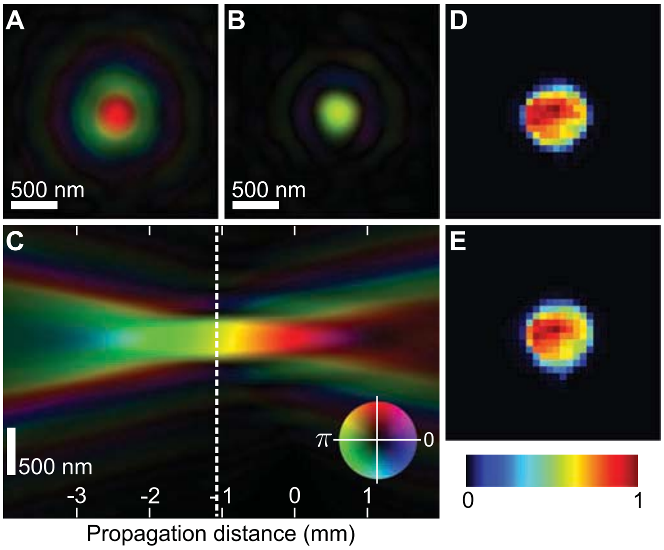


Fig. 4. Reconstructed probe. (**A**) Color rendition of the complex-valued probe at the specimen plane. (**B**) The computed wavefield propagated back to the focal plane, situated 1.2 mm upstream. (**C**) A section of the probe wavefield along the propagation direction, as calculated from the data in (A). The dashed line, indicating the plane of (B), is at the waist of the wave distribution. (**D**) The squared amplitude of the Fourier transform of the probe. (**E**) Measured diffraction pattern of the probe without a specimen. This data, which was not used in the reconstruction, is in good agreement with the retrieved probe.

specimen. The amplitude component carries the absorption information, and the phase expresses the phase shift experienced by a wave when traversing the specimen. In the present case, the phase difference caused by the 1.0-µm-thick gold structure in the zones is measured to be about 0.67 π, in very good agreement with the expected value of 0.65 π. We note that a complex reconstruction carries the necessary information to reproduce any traditional microscopy arrangement. For instance, a differential phase contrast image, similar to Fig. 2, C and D, is simply given by the gradient of Fig. 3B. An example of a simulated phase contrast image is shown in fig. S1C.

The wavefield of the probe, which is retrieved at the same time as the object, is shown in Fig. 4A. The knowledge of both the phase and the amplitude of the wave gives the most complete information for subsequent analysis. Computed propagation of this wavefield shows that the specimen was placed 1.2 mm behind the focal plane and was thus illuminated with a slightly curved and diverging wavefront.

Unlike the few other known methods (*11*, *28*), the probe reconstruction does not require any knowledge of the pupil of the Fresnel zone plate. In fact, it does not require a pupil plane. Probes formed with refractive (*29*) or reflective (*30*) optics can be equally well reconstructed, as well as unfocused probes formed with a mask. The door to a general wavefront characterization procedure appears to be even more widely open when one realizes the symmetry in the multiplicative relation between P and O in Eq. 1. Transferring the oversampling requirement to the object function placed downstream—for instance, by using a simple pinhole—allows the characterization of arbitrary large and extended regions of an incoming wavefield. In this wavefront-sensing configuration, what was the "probe" becomes the main object of interest: It can be the propagated wave perturbation from any specimen or optical device. This feature will be useful for the characterization of focusing devices to be used at future coherent x-ray light sources.

We have demonstrated an imaging technique that is fully compatible with STXM and has the high-resolution potential of CDI approaches. The reconstruction method removes the main difficulty of ptychography by retrieving the complex-valued image of both the specimen and the probe. The method is noninvasive, and radiation damage can be reduced by a combination of fast large-area scans with longer high-resolution imaging of regions of interest. The technique can be used with hard and soft x-rays, can be combined with other scanning methods such as fluorescence imaging, and can be extended to nanodiffraction mapping or nanospectroscopy. Higher coherent flux and improved focusing optics (*31*) should soon provide conditions for better-than-10-nm resolutions, making possible the imaging of the finest structures in state-of-the-art electronics devices or the macromolecular assemblies in organic tissues. Future developments comprise extension of the method to 3D reconstructions of inorganic and organic specimens as a possible answer to the current limitations of diffraction microscopy (*32*).

References and Notes

1. J. Miao, P. Charalambous, J. Kirz, D. Sayre, *Nature* **400**, 342 (1999).
2. H. N. Chapman *et al.*, *Nat. Phys.* **2**, 839 (2006).
3. H. N. Chapman *et al.*, *J. Opt. Soc. Am. A* **23**, 1179 (2006).
4. G. J. Williams, M. A. Pfeifer, I. A. Vartanyants, I. K. Robinson, *Phys. Rev. Lett.* **90**, 175501 (2003).
5. J. Miao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 110 (2003).
6. D. Shapiro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15343 (2005).
7. I. McNulty *et al.*, *Science* **256**, 1009 (1992).
8. S. Eisebitt *et al.*, *Nature* **432**, 885 (2004).
9. J. M. Rodenburg, R. H. T. Bates, *Philos. Trans. R. Soc. London Ser. A* **339**, 521 (1992).
10. P. Nellist, B. C. McCallum, J. M. Rodenburg, *Nature* **374**, 630 (1995).
11. H. N. Chapman, *Ultramicroscopy* **66**, 153 (1996).
12. J. M. Rodenburg, A. C. Hurst, A. G. Cullis, *Ultramicroscopy* **107**, 227 (2007).
13. J. M. Rodenburg *et al.*, *Phys. Rev. Lett.* **98**, 034801 (2007).
14. R. Hegerl, W. Hoppe, *Ber. Bunsenges. Phys. Chem* **74**, 1148 (1970).
15. G. Morrison, A. Gianoncelli, B. Kaulich, D. Bacescu, J. Kovac, in *Proceedings of the Eighth International Conference on X-ray Microscopy, IPAP Conference Series* **7**, 377 (2005).
16. B. Hornberger, M. Feser, C. Jacobsen, *Ultramicroscopy* **107**, 644 (2007).
17. C. Brönnimann *et al.*, *J. Synchrotron Radiat.* **13**, 120 (2006).
18. Materials and methods are available as supporting material on *Science* Online.
19. The validity of this relation, discussed in (*18*), depends on the thickness of the specimen, on the probe's depth of focus, and on the resolution of the reconstruction.
20. Measurements were carried at the coherent small-angle x-ray scattering (cSAXS) beamline of the Swiss Light Source at Paul Scherrer Institut.
21. At each point $\mathbf{r}_j$ of the raster grid, assuming that O varies slowly within the illuminated region, the measured diffraction pattern can be approximated with $I_j(\mathbf{q}) \approx |O(\mathbf{r}_j)|^2 |\tilde{P}[\mathbf{q} - \nabla\phi(\mathbf{r}_j)]|^2$, where $\tilde{P}$ is the Fourier transform of P. The specimen's transmission, $|O(\mathbf{r}_j)|^2$, is read from the overall attenuation of the signal, and the phase gradient, $\nabla\phi(\mathbf{r}_j)$, is seen as a position shift of the diffraction pattern on the detector.
22. J. M. Rodenburg, H. M. L. Faulkner, *Appl. Phys. Lett.* **85**, 4795 (2004).
23. H. M. L. Faulkner, J. M. Rodenburg, *Phys. Rev. Lett.* **93**, 023903 (2004).
24. V. Elser, *J. Opt. Soc. Am. A* **20**, 40 (2003).
25. V. Elser, I. Rankenburg, P. Thibault, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 418 (2007).
26. In the Fraunhoffer regime, the wave in the transverse plane at the exit of the specimen $\psi(\mathbf{r})$ is related to the intensity measurement by $I(\mathbf{q}) = |\tilde{\psi}(\mathbf{q})|^2$, where $\tilde{\psi}(\mathbf{q})$ is the 2D Fourier transform of $\psi(\mathbf{r})$. This equality constrains the amplitude of $\tilde{\psi}(\mathbf{q})$ but not its phase.
27. The best object and probe functions, given a set of views (ψ_1, ψ_2, ...) are the solution of the coupled system $O(\mathbf{r}) = \dfrac{\sum_j P^*(\mathbf{r} - \mathbf{r}_j)\psi_j(\mathbf{r})}{\sum_j |P(\mathbf{r} - \mathbf{r}_j)|^2}$, $P(\mathbf{r}) = \dfrac{\sum_j O^*(\mathbf{r} + \mathbf{r}_j)\psi_j(\mathbf{r} + \mathbf{r}_j)}{\sum_j |O(\mathbf{r} + \mathbf{r}_j)|^2}$. More details on the algorithm are provided in (*18*).
28. H. M. Quiney, A. G. Peele, Z. Cai, D. Paterson, K. A. Nugent, *Nat. Phys.* **2**, 101 (2006).
29. C. Schroer *et al.*, *Appl. Phys. Lett.* **82**, 1485 (2003).
30. H. Mimura *et al.*, *Appl. Phys. Lett.* **90**, 051903 (2007).
31. K. Jefimovs *et al.*, *Phys. Rev. Lett.* **99**, 264801 (2007).
32. D. Sayre, *Acta Crystallogr. A* **64**, 33 (2008).

33. We wish to thank P. Kraft, E. F. Eikenberry, B. Henrich, and C. Brönnimann for the Pilatus detector commissioning and the ongoing support; K. Jefimovs for the fabrication of nanostructures; K. Nygard for producing the scanning electron microscopy (SEM) picture of the specimen; X. Donath for the technical support at the cSAXS beamline; and J. Rodenburg for introducing us to ptychography. This work was performed at the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland. P.T. acknowledges financial support from the Fonds Québécois de la recherche sur la nature et les technologies (FQRNT).

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5887/379/DC1
Materials and Methods
Fig. S1
References

2 April 2008; accepted 13 June 2008
10.1126/science.1158573

Observations of Intergranular Stress Corrosion Cracking in a Grain-Mapped Polycrystal

A. King,[1,2]* G. Johnson,[1,2] D. Engelberg,[1] W. Ludwig,[2,3] J. Marrow[1]

Nondestructive three-dimensional mapping of grain shape, crystallographic orientation, and grain boundary geometry by diffraction contrast tomography (DCT) provides opportunities for the study of the interaction between intergranular stress corrosion cracking and microstructure. A stress corrosion crack was grown through a volume of sensitized austenitic stainless steel mapped with DCT and observed in situ by synchrotron tomography. Several sensitization-resistant crack-bridging boundaries were identified, and although they have special geometric properties, they are not the twin variant boundaries usually maximized during grain boundary engineering.

Intergranular stress corrosion cracking (IGSCC) is the progressive nucleation and growth of cracks by localized corrosion along the grain boundaries in metals in the presence of stress or strain. Crack growth is driven by the inelastic deformation at the sharp crack tip. Thus, IGSCC is the result of the combination of susceptible material, critical environment, and sufficient mechanical driving force. It is a critical failure mechanism in some components of a power generation plant (*1*), where the cracking of austenitic stainless steels can result from the sensitization of certain grain boundaries after heat treatment (such as post-weld stress relief) or fast neutron irradiation in a nuclear plant.

Sensitization decreases the local resistance of grain boundaries to corrosion. Thermal sensitization is due to chromium carbide precipitation at the grain boundaries (*2*). The slow diffusing chromium is depleted in the region adjacent to the carbides, to a degree sufficient to substantially reduce the oxidation resistance. Chromium depletion at grain boundaries also occurs with irradiation damage (*3*), although point defect creation and

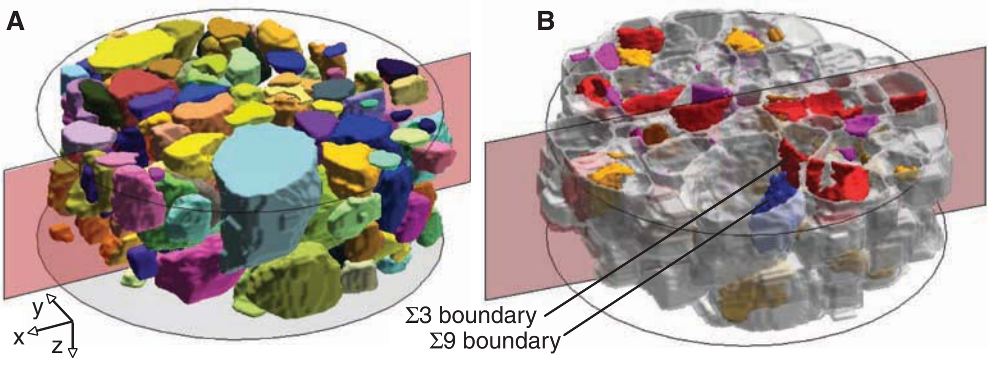


Fig. 1. Part of the 3D grain map obtained by DCT, including 169 grains (a total of 362 grains were mapped). The circumference of the sample is outlined, and the plane of the 2D section in Fig. 3 is also shown. (**A**) Grains colored using a RBG scale, according to their crystallographic orientation. (**B**) Low Σ CSL grain boundaries are shown in color: low-angle $\Sigma 1$, orange; $\Sigma 3$, red; $\Sigma 9$, blue; other boundaries $\Sigma \leq 29$, purple.

---

[1] School of Materials, University of Manchester, Manchester M13 9PL, UK. [2] European Synchrotron Radiation Facility, Grenoble, France. [3] Matériaux: Ingénierie et Science, Institut Nacional des Sciences Appliquées de Lyon, Villeurbanne, France.

*To whom correspondence should be addressed. E-mail: andrew.king@esrf.fr

changes in dislocation behavior may also be important factors (*4*).

Sensitization resistance is affected by grain boundary structure, which is commonly described by the coincident site lattice (CSL) notation. This notation describes the orientation relationship between the crystal lattices of grains adjacent to the grain boundary and is generally obtained using two-dimensional (2D) electron backscatter diffraction (EBSD) techniques (*5*). CSL notation provides no information on the actual grain boundary plane for most boundaries. The crystallographic habit plane of coherent Σ3 twins is well defined, for example, but not that of incoherent Σ3 twins. However, the grain boundary plane has an important effect on the grain boundary structure (*5*), and descriptions such as the fraction of Σ3 boundaries in the microstructure neglect this effect. Nonetheless, this description has been sufficient to enable the resistance to intergranular degradation of materials to be improved by grain boundary engineering (*6*). This is done typically by maximizing the fraction of twin variant grain boundaries ($\Sigma 3^n$, $1 \leq n \leq 3$, where n is = 1, 2, or 3), or more generally, low Σ CSL boundaries (≤Σ29).

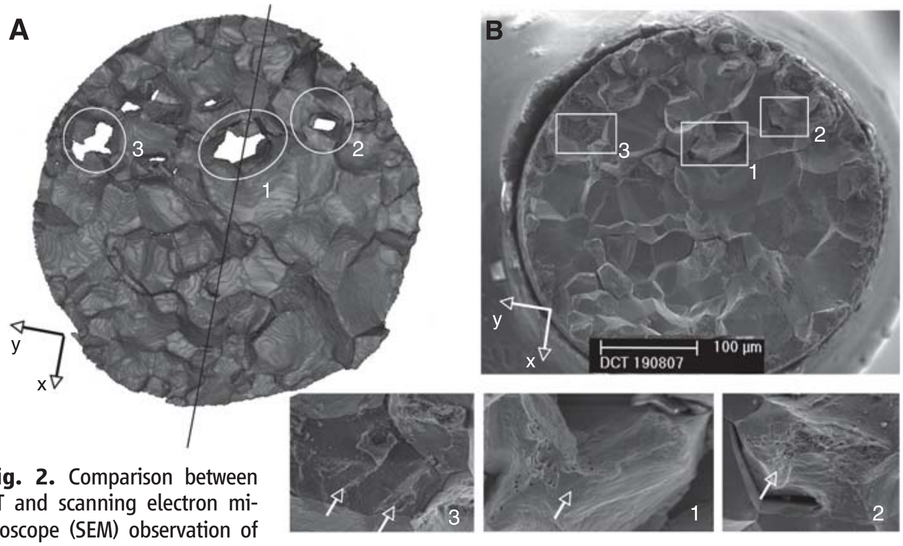


Fig. 2. Comparison between CT and scanning electron microscope (SEM) observation of the IGSCC fracture surface. (**A**) Surface rendering of CT data for the final crack final step, with three major bridging ligaments marked with numbers. The line denotes the plane of the 2D section in Fig. 3. (**B**) SEM image of the fracture surface after failure. Higher-resolution SEM observations (insets) show ductile features at the positions of the three holes in the CT data set.

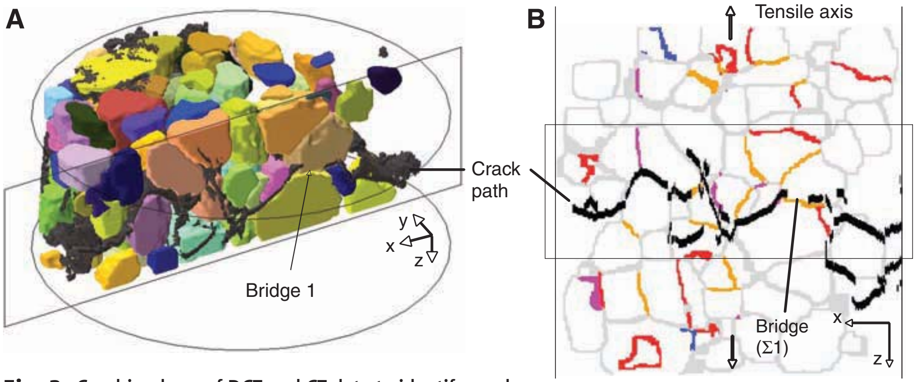
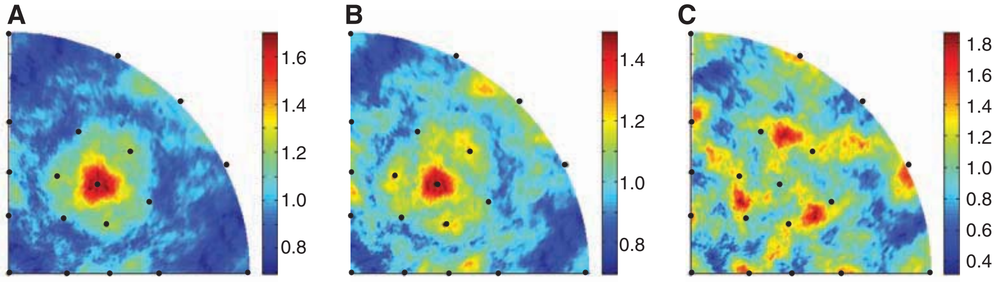


Fig. 3. Combined use of DCT and CT data to identify crack-bridging grain boundary structure. (**A**) Cracks obtained from CT data are shown in black, at the final step before sample failure, and compared with DCT data of 3D grain shapes. (**B**) 2D section of the grain boundaries, identified by DCT, compared with the crack path identified by CT. The boundaries are colored as in Fig. 1, and a crack bridge is indicated.

Fig. 4. Pole figures showing the frequency distribution of grain boundary planes in the studied sample. {hkl} poles are marked up to {221}. (**A**) All boundaries, (**B**) all boundaries excluding Σ3 twins, and (**C**) boundaries forming the fracture surface.



High-resolution synchrotron tomography has been applied previously to obtain 3D in situ observations of IGSCC, using sensitized austenitic stainless steel (*7*). Evidence for crack bridging by individual grain boundaries resistant to cracking was obtained; these nonsensitized, corrosion-resistant grain boundaries fail in a ductile manner at high strain. They are proposed to mechanically restrict the opening of the crack tip in a manner analogous to fibers in a composite material. Such crack bridging would reduce the inelastic crack tip strains in response to the applied stress and, consequently, would decrease the stress corrosion crack growth rate. On the basis of these observations, 3D modeling of intergranular stress corrosion crack nucleation predicts that grain boundary engineering to increase the proportion of sensitization-resistant boundaries substantially increases the incubation period for crack nuclei and thereby increases IGSCC resistance (*8*, *9*). Validation and further development of such IGSCC models—to address intergranular strains from grain misorientations (*10*), for example—requires observations of the 3D interaction between cracks and microstructure and, therefore, the full 3D characterization of the microstructure itself.

There are a number of approaches to 3D microstructure characterization. Destructive analysis of the microstructure is possible, although 2D techniques are not generally able to fully characterize a grain boundary. The trace of the intersection between the boundary plane and the surface can be identified, but the inclination of the boundary plane is not known. Thus, only four of the five parameters required for a full grain boundary description can be obtained (*5*). Statistical 3D boundary descriptions may be obtained from 2D data with the use of stereological methods (*11*), but they are unsuitable for studies of local events such as crack behavior. Destructive serial-sectioning for 3D characterization of local features is also possible (*12*). However, such studies are time-consuming, do not necessarily provide data on all grain boundaries, and are difficult to apply to damaged samples that contain cracks. Recently, two nondestructive 3D orientation mapping techniques employing synchrotron radiation have become available, known as differential aperture x-ray microscopy (*13*) and 3D x-ray diffraction microscopy (3DXRD) (*14*), respectively. The technique used in this work, called diffraction contrast tomography (DCT), is closely related to 3DXRD and has been developed by some of the authors in collaboration with the Risø group.

DCT currently allows the characterization of plastically nondeformed, polycrystalline materials in terms of 3D grain shapes and crystallographic orientations and simultaneously provides a picture of the material's microstructure using x-ray absorption contrast. Each voxel (*15*) is assigned to a grain within the characterized volume, and the attenuation, as well as the crystal orientation, of each voxel is known. The principles and analysis methods of DCT are described in (*16*, *17*) and are briefly summarized together with the use of synchrotron computed tomography (CT) to image the development of the crack in (*18*). To study the interaction of the stress corrosion crack with the microstructure, it is necessary to marry the DCT and CT observations, between which there will be rotational and translational differences.

The sample used was an austenitic stainless steel wire, heat treated to produce a fully recrystallized microstructure with a mean grain size of ~50 μm and then thermally sensitized for IGSCC. The sample preparation is described in more detail in (*18*). The DCT data set provides a 3D map (Fig. 1A) of the grain shapes and their crystal orientations. The map comprises 362 grains, giving some 1600 fully characterized grain boundaries. For visualization, the grain surfaces are rendered and colored according to their crystallographic orientation, using the three components of the Rodrigues's vector (*19*, *20*) to obtain RBG (red/blue/green) values with appropriate scaling. The grains' shapes are observed to be self-consistent with their neighbors. The narrow gaps (typically 15 to 30 μm) between the grains are an artifact of the data processing and may be removed in a simple post-processing dilation operation. Before this post-processing, the space-filling of the grain map is 59%. The grain boundaries can be described with the use of the relative crystal lattice misorientation between the adjacent grains. This has been done according to the CSL convention, using Brandon's deviation criterion (*5*). In the 3D grain boundary map (Fig. 1B), the low-angle boundaries ($\Sigma1$), twins ($\Sigma3$), and low Σ CSL boundaries ($\Sigma5$ to $\Sigma29$) have been highlighted.

After the DCT characterization, the sample was loaded in a corrosive environment to initiate a stress corrosion crack, the propagation of which was studied in situ using synchrotron CT (*7*, *18*). Multiple initiation sites were observed, from which a dominant crack coalesced to propagate across the sample. The results presented here are for the final observation before sample failure. A 3D surface rendering of the crack at this stage is shown in Fig. 2A. Three holes in this surface can be readily identified. These are regions where the surrounding material is continuous across the crack, and they are therefore ligaments that have been isolated as the crack propagated around them. The sample was then allowed to fail by further SCC, and the final fracture surface, observed by scanning electron microscopy, is shown in Fig. 2B. The higher-resolution images (insets in Fig. 2B) at the location of the holes show ductile fracture surfaces, in contrast to the intergranular brittle character of the majority of the surface that is characteristic of IGSCC. The ductile regions are caused by the tearing to failure of the crack-bridging ligaments after the final tomography scan. Such regions are distributed over the entire fracture surface, indicating that they also develop during crack growth (*7*).

A rendering of the fracture surface from Fig. 2A is displayed in Fig. 3A, showing its alignment with the 3D grain map. The volume is sectioned at the position of one of the holes, or bridging ligaments, in the intergranular crack. As the DCT and CT data sets are aligned, each facet of the intergranular crack can be associated with a particular grain boundary. Hence, the crack can be described in terms of the structure and orientation of the grain boundaries along which it propagates. This combined technique also allows for the identification of the grain boundary at the position of each of the three holes, or crack bridges, shown in Fig. 2A.

An example (crack bridge 1) is shown in Fig. 3B, with the trace of the crack on a map of the grain boundaries, colored according to their CSL designation. The hole in the CT image of the crack coincides with a low-angle grain boundary (the misorientation angle between the crystal lattices of the grains is 9.4° ± 0.05°). The grain boundary normal, obtained by fitting a plane between the grains, was inclined at 19.6° ± 10° to the tensile stress axis. The uncertainty is largely due to the gaps between the grains in the DCT data set and to the curvature of grain boundary planes. Similar analysis shows that the hole at crack bridge 2 is located at a $\Sigma11$ boundary. Its boundary normal is at 32.2° ± 10° to the tensile axis. The grain boundary that coincides with the hole at crack bridge 3 is neither a low-angle boundary nor a low Σ CSL boundary. It has a crystal misorientation angle of 44.3° and a misorientation axis that is 1.1° ± 0.05° from <310>. This boundary has strong tilt character, with the misorientation axis lying at 86.8° ± 10° to the grain boundary normal. The grain boundary plane lies within 3.9° ± 10° of the {001} pole in one of the two adjoining grains and within 5.0° ± 10° of the {123} pole in the other, indicating that it has an orientation close to a low {hkl} index plane in both grains.

Low-angle and low Σ CSL grain boundaries are known to generally have special properties. Similarly, it is proposed that boundaries close to low index planes may be beneficial, despite not being low Σ CSL (*21*). Because of the 3D characterization of the sample, we are able to determine the approximate plane of all grain boundaries. The frequency distribution of grain boundary plane orientations obtained from the fracture surface and the bulk microstructure is shown in Fig. 4. The bulk microstructure is shown, both including and excluding the $\Sigma3$ twins. The number of boundaries on the fracture surface that are close to low {hkl} index planes such as {111} is low when compared with the bulk microstructure, although the fracture surface will include some nonsensitized boundaries that failed in a ductile manner.

These in situ observations are direct confirmation of the resistance of individual grain boundaries to IGSCC and their role in crack-bridge development. This finding is consistent with reports that grain boundary engineering to reduce the proportion of sensitized grain boundaries can improve IGSCC resistance (*5*, *6*, *9*). Furthermore, the results suggest that a specific range of grain boundaries have "special" properties and may be involved in crack bridging during SCC, beyond the twin variant grain boundaries ($\Sigma3^n$) usually targeted in grain boundary engineering.

Knowledge of the behavior of individual grain boundaries is most beneficial to the study of intergranular failure. The application of 3D orientation mapping techniques, such as DCT, may provide local and spatial information on the connectivity and clustering of certain grain boundary structures throughout the bulk material. This more complete description of microstructure, together with an understanding of the structure of individual boundaries and their cracking mechanisms [obtained by transmission electron microscopy studies, for example (*22*, *23*)], is the foundation for more realistic models of crack behavior. The combined approach described here has potential applications to other forms of degradation, including fatigue and transgranular SCC.

References and Notes

1. P. M. Scott, *Corrosion* **56**, 771 (2000).
2. S. M. Bruemmer, G. S. Was, *J. Nucl. Mater.* **216**, 348 (1994).
3. P. Scott, *J. Nucl. Mater.* **211**, 101 (1994).
4. J. T. Busby, G. S. Was, E. A. Kenik, *J. Nucl. Mater.* **302**, 20 (2002).
5. V. Randle, *Acta Mater.* **46**, 1459 (1998).
6. E. M. Lehockey, A. M. Brennenstuhl, I. Thompson, *Corros. Sci.* **46**, 2383 (2004).
7. L. Babout, T. J. Marrow, D. Engelberg, P. J. Withers, *Mater. Sci. Technol.* **22**, 1068 (2006).
8. A. P. Jivkov, N. P. C. Stevens, T. J. Marrow, *Comput. Mater. Sci.* **38**, 442 (2006).
9. A. P. Jivkov, N. P. C. Stevens, T. J. Marrow, *Theor. Appl. Fract. Mech.* **48**, 187 (2007).
10. O. Diard, S. Leclercq, G. Rousselier, G. Cailletaud, *Comput. Mater. Sci.* **25**, 73 (2002).
11. V. Randle, *J. Microsc.* **222**, 69 (2006).
12. A. C. Lewis *et al.*, *Mater. Sci. Eng. A* **418**, 11 (2006).
13. B. C. Larson, W. Yang, G. E. Ice, J. D. Budai, J. Z. Tischler, *Nature* **415**, 887 (2002).
14. H. F. Poulsen, *Three-Dimensional X-Ray Diffraction Microscopy: Mapping Polycrystals and Their Dynamics* (Springer, Berlin, 2004).
15. A voxel in the 3D data set is the analog of a pixel in a 2D digital image, such as an EBSD map.
16. W. Ludwig, S. Schmidt, E. M. Lauridsen, H. F. Poulsen, *J. Appl. Crystallogr.* **41**, 302 (2008).
17. G. Johnson, A. King, M. Gonzalves Honnicke, T. J. Marrow, W. Ludwig, *J. Appl. Crystallogr.* **41**, 310 (2008).
18. Materials and methods are available as supporting material on *Science* Online.
19. O. Rodrigues, *J. Math. Pures. Appl.* **5**, 380 (1840).
20. F. C. Frank, *Metall. Trans. A* **19**, 403 (1988).

21. V. Randle, *Scr. Mater.* **54**, 1011 (2006).
22. S. R. Ortner, *Acta Metall. Mater.* **39**, 341 (1991).
23. Y. Z. Huang, J. M. Titchmarsh, *Acta Mater.* **54**, 635 (2006).
24. A.K. and G.J. wish to acknowledge funding received from the Engineering and Physical Sciences Research Council (U.K.). W.L. also acknowledges additional support received from the Danish National Research Foundation.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5887/382/DC1
Materials and Methods
Fig. S1
References

7 February 2008; accepted 12 June 2008
10.1126/science.1156211

Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene

Changgu Lee,[1,2] Xiaoding Wei,[1] Jeffrey W. Kysar,[1,3] James Hone[1,2,4]*

We measured the elastic properties and intrinsic breaking strength of free-standing monolayer graphene membranes by nanoindentation in an atomic force microscope. The force-displacement behavior is interpreted within a framework of nonlinear elastic stress-strain response, and yields second- and third-order elastic stiffnesses of 340 newtons per meter ($N\ m^{-1}$) and $-690\ N\ m^{-1}$, respectively. The breaking strength is 42 $N\ m^{-1}$ and represents the intrinsic strength of a defect-free sheet. These quantities correspond to a Young's modulus of $E = 1.0$ terapascals, third-order elastic stiffness of $D = -2.0$ terapascals, and intrinsic strength of $\sigma_{int} = 130$ gigapascals for bulk graphite. These experiments establish graphene as the strongest material ever measured, and show that atomically perfect nanoscale materials can be mechanically tested to deformations well beyond the linear regime.

In 1921, Griffith published a groundbreaking study on the fracture of brittle materials that established the relationship between the change of potential energy of a brittle system with crack growth and the free energy of a newly created surface (*1*). As a result of this insight, Griffith deduced that the actual breaking strength of a brittle material is governed by the sizes of defects and flaws within the material, rather than the intrinsic strength of its atomic bonds. To emphasize the point, Griffith wrote that "in the limit, in fact, a fiber consisting of a single line of molecules must possess the theoretical molecular tensile strength," the maximum stress that can be supported by the material prior to failure in a pristine material without defects, here denoted as the intrinsic strength. He then proceeded to experimentally estimate the intrinsic tensile strength by measuring the breaking strength of a series of glass fibers with progressively smaller diameters and extrapolating the results to an atomic radius. He extrapolated an intrinsic strength of about $E/9$, where E is the elastic stiffness (Young's modulus) of the material under uniaxial tension. The concepts related to fracture have been well developed in the intervening decades; however, a direct and repeatable measurement of the intrinsic breaking strength of a material has remained elusive. We probed the intrinsic strength of monolayer graphene, as well as its linear and nonlinear elastic properties.

Graphene, which consists of a two-dimensional (2D) sheet of covalently bonded carbon atoms, forms the basis of both 3D graphite and 1D carbon nanotubes. Its intrinsic strength, predicted to exceed that of any other material (*2*), motivates the use of carbon-fiber reinforcements in advanced composites, and may permit such exotic structures as a "space elevator" if macroscopic fibers close to the theoretical strength can be realized. However, the intrinsic strength of this material has still not been definitively measured because of the inevitable presence of defects and grain boundaries in macroscopic samples. In the past few years, multiple studies (*3*–*10*) of carbon nanotubes have confirmed their high stiffness and strength. However, determination of these quantities has been difficult because of uncertainty in the sample geometry, stress concentration at clamping points, structural defects, and unknown load distribution among shells in multiwalled nanotubes. Recent experimental advances (*11*) now permit the study of individual graphene sheets. We used atomic force microscope (AFM) nanoindentation to measure the mechanical properties of monolayer graphene membranes suspended over open holes. This technique has recently been used to study multilayer graphene (*12*, *13*) and offers three important advantages over experiments on nanotubes: The sample geometry can be precisely defined, the 2D structure is less sensitive to the presence of a single defect, and the sheet is clamped around the entire hole circumference, as opposed to two points in the case of nanotubes.

For this study, a 5-by-5-mm array of circular wells (diameters 1.5 μm and 1 μm, depth 500 nm) was patterned onto a Si substrate with a 300-nm SiO_2 epilayer by nanoimprint lithography and reactive ion etching (Fig. 1). Graphite flakes were then mechanically deposited onto the substrate (*14*). Optical microscopy was used to find flakes of monolayer graphene, whose thicknesses were confirmed with Raman spectroscopy (*15*) (fig. S1). Figure 1A shows a monolayer flake deposited over many circular wells to form a series of free-standing membranes. Noncontact mode AFM imaging (Fig. 1B) confirmed that the

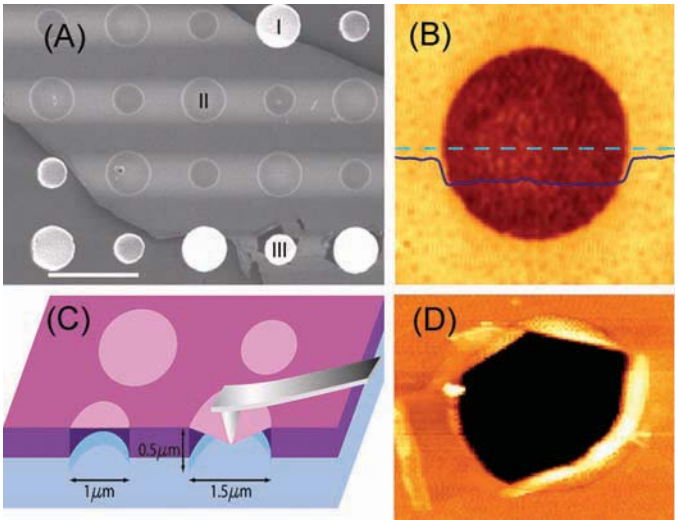


Fig. 1. Images of suspended graphene membranes. (**A**) Scanning electron micrograph of a large graphene flake spanning an array of circular holes 1 μm and 1.5 μm in diameter. Area I shows a hole partially covered by graphene, area II is fully covered, and area III is fractured from indentation. Scale bar, 3 μm. (**B**) Noncontact mode AFM image of one membrane, 1.5 μm in diameter. The solid blue line is a height profile along the dashed line. The step height at the edge of the membrane is about 2.5 nm. (**C**) Schematic of nanoindentation on suspended graphene membrane. (**D**) AFM image of a fractured membrane.

---

[1] Department of Mechanical Engineering, Columbia University, New York, NY 10027, USA. [2] Defense Advanced Research Projects Agency Center for Integrated Micro/Nano-Electromechanical Transducers (iMINT), Columbia University, New York, NY 10027, USA. [3] Center for Nanostructured Materials, Columbia University, New York, NY 10027, USA. [4] Center for Electronic Transport in Molecular Nanostructures, Columbia University, New York, NY 10027, USA.

*To whom correspondence should be addressed. E-mail: jh2228@columbia.edu

graphene membranes were stretched tautly across the well openings. It also revealed that the graphene adheres to the vertical wall of the hole for 2 to ~10 nm, presumably because of van der Waals attraction to the substrate.

The mechanical properties of the free-standing films were probed by indenting the center of each film with an AFM (XE-100, Park Systems), as illustrated in Fig. 1C. Because of the strength of the films, cantilevers with diamond tips were used for this study. Two different cantilevers were used, with tip radii of 27.5 and 16.5 nm, as measured before and after indentation via transmission electron micrography (TEM) (fig. S2). The corresponding cantilever spring constants were calibrated against a reference cantilever (*16*). Prior to indenting, the graphene membranes were scanned in a noncontact AFM mode and the AFM tip was subsequently positioned to within 50 nm of the center. Mechanical testing was performed at a constant displacement rate, followed by load reversal. This cycle was repeated several times for each film tested. The data showed no hysteresis, which demonstrated the elastic behavior of the film and showed that the graphene film did not slip around the periphery of the well. The force-displacement measurements were highly repeatable; data sets from different flakes, membrane diameters, displacement rates, and indenter tip radii yielded values of the elastic modulus that were statistically indistinguishable. Once the data for elastic properties of the films were recorded, the films were once again indented at the same rate, but this time to failure. The force-displacement data were processed to determine the elastic properties and breaking stress of the graphene membranes. Full details of the experimental procedure and data analysis can be found in (*16*).

The elastic response of the graphene must be considered nonlinear because the stress-strain response must curve over to a maximum point that defines the intrinsic breaking stress. An elastic (i.e., reversible) response implies the existence of an energy potential that is a function of strain that can be expressed as a Taylor series in powers of strain. The lowest-order (quadratic) term leads to linear elastic response. The third-order term gives rise to nonlinear elastic behavior (*17*). The resulting isotropic elastic response under uniaxial extension can be expressed as

$$\sigma = E\varepsilon + D\varepsilon^2 \qquad (1)$$

where σ is the symmetric second Piola-Kirchhoff stress, ε is the uniaxial Lagrangian strain, E is the Young's modulus, and D is the third-order elastic modulus. The value of D is typically negative, so the presence of the second-order term leads to a lessening of stiffness at high tensile strains and an increasingly stiff response at high compressive strains. E was determined from components of the second-order fourth-rank stiffness tensor (with two independent components for an isotropic material), whereas D was determined from components of both the second-order fourth-rank stiffness tensor and the third-order sixth-rank stiffness tensor (with three independent components for an isotropic material). Numerical simulations of graphene sheets and nanotubes suggest that a nonlinear elastic response is appropriate (*18*, *19*).

This thermodynamically rigorous nonlinear form of the stress-strain response should capture the salient features of the elastic behavior of graphene. The maximum of the elastic stress-strain response defines the intrinsic stress, which for this functional form is $\sigma_{\text{int}} = -E^2/4D$ at the strain $\varepsilon_{\text{int}} = -E/2D$, so it remains only to determine E and D from the experimental results. We determined the value of E based on the experimental force-displacement data and inferred the value of D from the experimental breaking force.

Monolayer graphene is a true 2D material, so its strain energy density is normalized by the area of the graphene sheet rather than by the volume. Therefore, its behavior under tensile loading is properly described by a 2D stress σ^{2D} and elastic constants E^{2D} and D^{2D} with units of force/length. The equations given above are valid for any dimensionality. For purposes of comparison to bulk graphite and other materials, these quantities can be divided by the interlayer spacing in graphite [$h = 0.335$ nm (*20*)] in order to obtain the corresponding 3D parameters. However, the derived quantities are not intrinsic properties of the single sheet and cannot be used to predict other mechanical properties, such as bending stiffness (*21*).

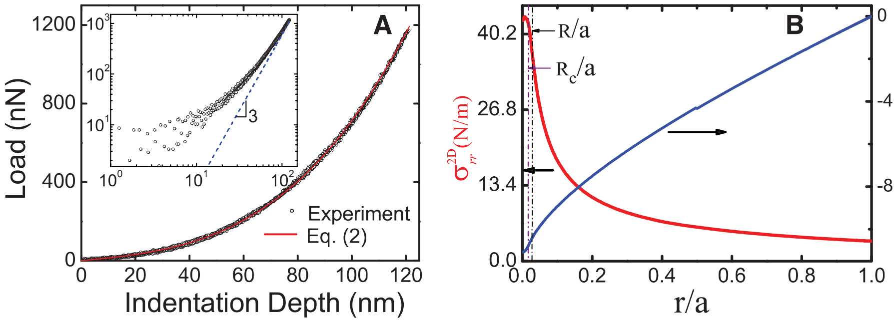


Fig. 2. (**A**) Loading/unloading curve and curve fitting to Eq. 2. The curve approaches cubic behavior at high loads (inset). (**B**) Maximum stress and deflection of graphene membrane versus normalized radial distance at maximum loading (simulation based on nonlinear elastic behavior in Eq. 1). The dashed lines indicate the tip radius R and contact radius R_c.

Even at the maximum curvatures present in these experiments, the energy from bending the graphene membrane is three orders of magnitude smaller than the energy from in-plane strain, using ab initio values for the in-plane stiffness and flexural rigidity (*22*). Therefore, the graphene can be modeled as a 2D membrane (i.e., it has zero bending stiffness). We performed detailed finite-element analyses for this geometry and loading based on the constitutive relations in Eq. 1, using various values of D^{2D}, tip radius, and indenter position (*16*). Three important points emerged from this analysis. First, the resulting force-displacement curve made by using the nonlinear elastic model was virtually indistinguishable from that of the linear model (where $D^{2D} = 0$). The calculated radial stress distribution and shape of the deformed film (Fig. 2B) show why nonlinear effects can be ignored while simulating the force-displacement response: Even close to breaking, at most 1% of the graphene film is strained to the point where the nonlinear term in Eq. 1 becomes important. Second, the simulations demonstrated that the force-displacement curve is insensitive to the indenter tip radius when $R \ll a$, where R is the tip radius and a is the membrane diameter. Third, the force-displacement response is insensitive to the position of the tip (to within experimental uncertainty) if the indenter tip is located within $a/10$ of the center of the film. Therefore, for the purposes of modeling the force-displacement behavior, the system can be approximated as a clamped circular membrane, made of a linear isotropic elastic material, under central point loading. Isotropic mechanical properties were used because the graphene atomic lattice has sixfold rotation symmetry (*23*).

Given the above model, the force-displacement behavior can be approximated (*16*, *24*, *25*) as

$$F = \sigma_0^{2D}(\pi a)\left(\frac{\delta}{a}\right) + E^{2D}(q^3 a)\left(\frac{\delta}{a}\right)^3 \qquad (2)$$

where F is applied force, δ is the deflection at the center point, σ_0^{2D} is the pretension in the film, ν is Poisson's ratio [taken here as 0.165, the Poisson's ratio for graphite in the basal plane (*26*)], and $q = 1/(1.05 - 0.15\nu - 0.16\nu^2) = 1.02$ is a dimensionless constant. The solid line in Fig. 2A shows the least-squares curve fit of one set of experimental data, based on Eq. 2, taking σ_0^{2D} and E^{2D} as free parameters. The closeness of the fit further validates the appropriateness of this model.

For this study, membranes from two separate graphene flakes were tested. These included seven with 1-μm diameters and six with 1.5-μm diameters from flake 1, and five of each size from flake 2. Each flake was probed using a different cantilever. For each membrane, approximately three loading/unloading curves were collected, to multiple depths (between 20 and 100 nm). In total, 67 values of σ_0^{2D} and E^{2D} were determined by fitting to Eq. 2. Figure 3A shows the distribution of the derived values of E^{2D}. The mean value of E^{2D} is 342 N m^{-1}, with a standard deviation of 30 N m^{-1}. An estimate of the overall uncertainty in the measured value is discussed in (*16*). The elastic modulus values obtained from all data subsets (different flakes, tips, well radii, indentation depth, and indentation speed) were statistically indistinguishable.

Figure 3B shows the derived values for the film pretension, which range from 0.07 to 0.74 N m^{-1}, with flake 2 showing larger pretension on average. These values are remarkably high, higher than the fracture strengths of many conventional materials. We also observed that, because the membrane follows the wall for 2 to 10 nm, the actual membrane profile is ~0.2 to 1% longer than the nominal well diameter. This elongation would lead to an added stress of 0.7 to 3.4 N m^{-1}, larger than the measured pretension in most cases. Therefore, we conclude that the graphene is nominally in a compressed state when not suspended.

The intrinsic strength of graphene was measured by loading the membranes to the breaking point. Figure 4A shows four typical breaking curves, for both tip radii and well diameters. There is no sign of slippage or other irreversible deformation prior to catastrophic failure. Figure 1D shows an AFM image of a fractured membrane. The graphene film still hangs around the edge of the hole, which suggests that fracture started at the indentation point. The films break at large deflections (above 100 nm) and forces of about 1.8 μN and 2.9 μN for the smaller and larger indenter tips, respectively. These forces were large enough to break standard Si AFM tips; TEM inspection confirmed that the diamond tips used in this study were not damaged. From those four typical curves, a clear pattern emerges: The force-displacement behavior is insensitive to tip radius, but the breaking force is mainly a function of tip radius and shows no dependence on membrane size, because of the extreme stress concentration (see also Fig. 2B) under the indenter tip.

The maximum stress for a clamped, linear elastic, circular membrane under a spherical indenter as a function of applied load has been derived on the basis of a continuum model as

$$\sigma_m^{2D} = \left(\frac{FE^{2D}}{4\pi R}\right)^{\frac{1}{2}} \qquad (3)$$

where σ_m^{2D} is the maximum stress at the central point of the film (*27*). This analytical solution helps illuminate the relationship among breaking force, tip radius, and maximum stress, in particular showing that the breaking force should vary inversely with tip radius. Analyzing all of the measured data using Eq. 3 yields an average breaking strength of 55 N m^{-1}. However, because the model ignores nonlinear elasticity, this value overestimates the strength.

The breaking forces we measured in the graphene films strongly suggest that the films in the neighborhood of the tip are free of defects, so that the maximum stress in the film represents the intrinsic strength. Two observations support this argument. First, the magnitude of the stresses under the indenter tip that we observed is consistent with predictions of intrinsic strength based on ab initio calculations for monolayer graphene (*18*, *19*). Second, the distribution of breaking forces, as shown in Fig. 4B, was relatively narrow, whereas for brittle fracture one would expect a wider distribution depending on the random size, number, and position of defects under the tip. The second argument can be quantified by fitting the histograms to the Weibull distribution (*28*), which characterizes the failure of brittle materials with random defects (fig. S8). The statistics suggest that the material is defect-free, at least under the indenter tip. This is consistent with scanning tunneling microscopy measurements (*29*) of graphene films (from the same source material) that show no defects over regions of hundreds of square nanometers. Therefore, we used a nonlinear elastic model to treat the breaking force data as corresponding to the intrinsic strength of the graphene.

A series of numerical simulations was performed for the 1-μm-diameter graphene film and the 16.5-nm-radius indenter tip to determine the relationship between the indentation breaking force and the third-order elastic constant, D^{2D}. In each simulation, the breaking force was determined as that load at which the solution could no longer converge to an equilibrium state under the indenter tip because of the negative slope of the elastic response at strains larger than ε_{int}. The mean experimentally determined breaking force of 1770 nN was consistent with a value of D^{2D} = –690 N m^{-1}. This value of D^{2D} was then used in the numerical model for the 1-μm-diameter graphene film and the 27.5-nm-radius indenter. The simulation predicted a breaking force (2880 nN) virtually identical to the mean experimental value of 2890 nN. Therefore, the experimentally determined values of the second-order and third-order elastic stiffnesses for monolayer graphene are E^{2D} = 340 ± 50 N m^{-1}

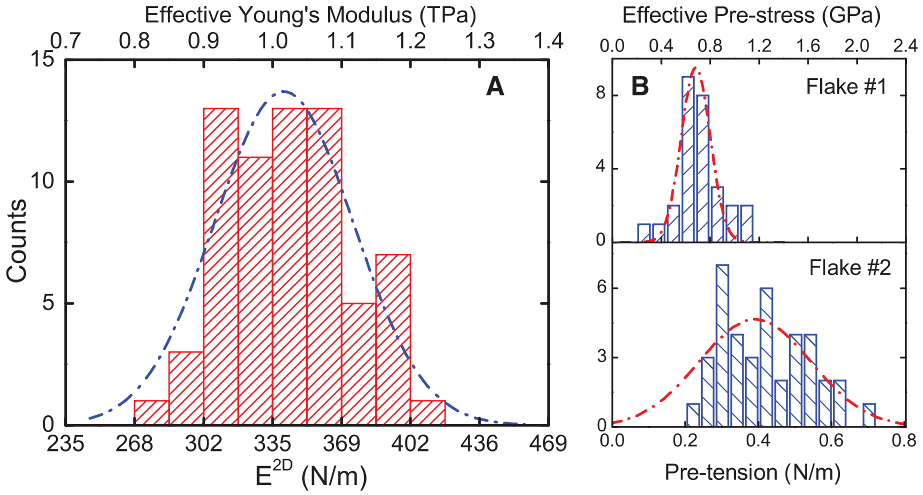


Fig. 3. Elastic response test results. (**A**) Histogram of elastic stiffness. (**B**) Histogram of film pretensions. Dashed lines in both plots represent Gaussian fits to data. The effective Young's modulus and prestress were obtained by dividing by the graphite interlayer spacing.

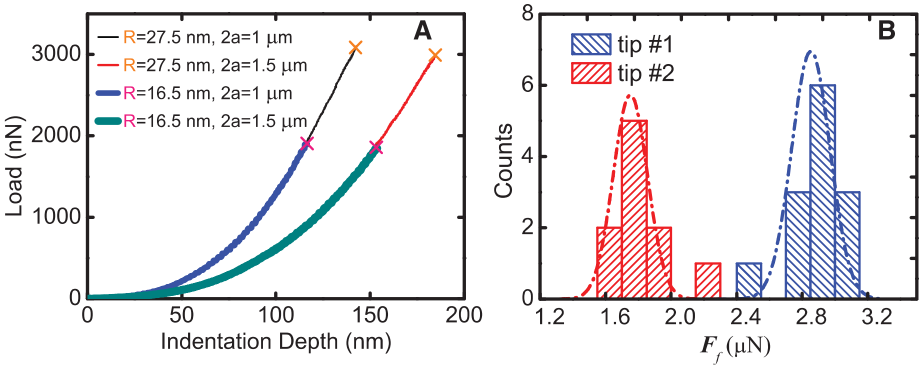


Fig. 4. Fracture test results. (**A**) Four typical tests, with different tip radii and film diameters; fracture loads are indicated by × marks. Breaking force depended strongly on tip radius but not on sample diameter. (**B**) Histogram and Gaussian distribution of breaking force for both tips.

and $D^{\mathrm{2D}} = -690 \pm 120$ N m^{-1}, respectively. The intrinsic strength is $\sigma_{\mathrm{int}}^{\mathrm{2D}} = 42 \pm 4$ N m^{-1}. These correspond to Young's modulus of $E = 1.0 \pm 0.1$ TPa and a third-order elastic stiffness of $D = -2.0 \pm 0.4$ TPa, assuming an effective graphene thickness of 0.335 nm. The corresponding intrinsic stress is $\sigma_{\mathrm{int}} = 130 \pm 10$ GPa at a strain of $\varepsilon_{\mathrm{int}} = 0.25$. Also, $\sigma_{\mathrm{int}} \approx E/8$, which is very close to the value obtained by Griffith.

We compared our measured values with those from other experiments. Experiments on bulk graphite (*26*) yield 1.02 ± 0.03 TPa for the in-plane Young's modulus. Tensile tests (*6*–*8*) have been reported for both single-walled carbon nanotubes and multi-walled carbon nanotubes (MWCNTs). A broad range of stiffness values (0.27 TPa to 1.47 TPa) was obtained, with breaking strengths ranging from 3.6 to 63 GPa and failure strain up to 12%. An in situ tensile test (*9*) in a TEM reported intrinsic strength of 150 ± 45 GPa for a defect-free MWCNT. At least two recent experiments report mechanical tests performed on suspended graphene films; nanoindentation of suspended multilayer graphene flakes has been used to measure their bending stiffness (*12*) and extract a Young's modulus of 0.5 TPa (*13*). The elastic deformation and failure strength of graphene have been simulated using ab initio methods (*18*, *19*). The predicted elastic response is highly nonlinear for strains above 10% and exhibits a maximum for pristine graphite between 118 and 121 GPa.

The intrinsic strength reported here serves as a benchmark for structural and mechanical applications, although the strength of macroscopic graphitic materials is still limited by the presence of defects and grain boundaries. In addition, these measurements demonstrate that third-order elastic constants associated with nonlinear behavior can be measured in atomically perfect nanoscale materials. These measurements can be used to validate models of atomic potentials far beyond the linear regime.

References and Notes

1. A. A. Griffith, *Philos. Trans. R. Soc. London Ser. A* **221**, 163 (1921).
2. Q. Zhao, M. B. Nardelli, J. Bernholc, *Phys. Rev. B* **65**, 144105 (2002).
3. A. Krishnan, E. Dujardin, T. W. Ebbesen, P. N. Yianilos, M. M. J. Treacy, *Phys. Rev. B* **58**, 14013 (1998).
4. J. P. Salvetat *et al.*, *Phys. Rev. Lett.* **82**, 944 (1999).
5. T. W. Tombler *et al.*, *Nature* **405**, 769 (2000).
6. S. Xie, W. Li, Z. Pan, B. Chang, L. Sun, *J. Phys. Chem. Solids* **61**, 1153 (2000).
7. M. F. Yu, B. S. Files, S. Arepalli, R. S. Ruoff, *Phys. Rev. Lett.* **84**, 5552 (2000).
8. M. F. Yu *et al.*, *Science* **287**, 637 (2000).
9. B. G. Demczyk *et al.*, *Mater. Sci. Eng. A* **334**, 173 (2002).
10. B. I. Yakobson, P. Avouris, in *Carbon Nanotubes* (Springer-Verlag, Berlin, 2001), vol. 80, pp. 287–327.
11. K. S. Novoselov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10451 (2005).
12. M. Poot, H. S. J. v. d. Zant, *Appl. Phys. Lett.* **92**, 063111 (2008).
13. I. W. Frank, D. M. Tanenbaum, A. M. Van der Zande, P. L. McEuen, *J. Vac. Sci. Technol. B* **25**, 2558 (2007).
14. K. S. Novoselov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10451 (2005).
15. A. C. Ferrari *et al.*, *Phys. Rev. Lett.* **97**, 187401 (2006).
16. Materials and methods are available as supporting material on *Science* Online.
17. R. N. Thurston, K. Brugger, *Phys. Rev.* **133**, A1604 (1964).
18. F. Liu, P. M. Ming, J. Li, *Phys. Rev. B* **76**, 064120 (2007).
19. R. Khare *et al.*, *Phys. Rev. B* **75**, 075412 (2007).
20. R. Al-Jishi, G. Dresselhaus, *Phys. Rev. B* **26**, 4514 (1982).
21. Y. Huang, J. Wu, K. C. Hwang, *Phys. Rev. B* **74**, 245413 (2006).
22. B. I. Yakobson, C. J. Brabec, J. Bernholc, *Phys. Rev. Lett.* **76**, 2511 (1996).
23. J. F. Nye, *Physical Properties of Crystals* (Oxford Univ. Press, New York, 1985).
24. K. T. Wan, S. Guo, D. A. Dillard, *Thin Solid Films* **425**, 150 (2003).
25. U. Komaragiri, M. R. Begley, *J. Appl. Mech.* **72**, 203 (2005).
26. O. L. Blakslee, *J. Appl. Phys.* **41**, 3373 (1970).
27. N. M. Bhatia, W. Nachbar, *Int. J. Non-Linear Mech.* **3**, 307 (1968).
28. B. Lawn, *Fracture of Brittle Solids* (Cambridge Univ. Press, New York, ed. 2, 1993).
29. E. Stolyarova *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9209 (2007).
30. We thank S. Berciaud for Raman spectroscopy of the graphene samples, Namiki Inc. for AFM cantilevers, and J. Hay and W. Oliver (MTS Nano Instruments) for useful interactions. We acknowledge support from NSF under awards CHE-0117752, CMMI-0500239, and DMR-0650555; iMINT (Award HR0011-06-1-0048); Air Force Office of Scientific Research grant FA9550-06-1-0214; and New York State Office of Science, Technology, and Academic Research. This work used shared experimental facilities supported primarily by the Materials Research Science and Engineering Center Program of NSF under award DMR-0213574, and the Cornell Nanoscale Science and Technology Facility, a member of the National Nanotechnology Infrastructure Network, which is supported by NSF (ECS-0335765).

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5887/385/DC1
Materials and Methods
Figs. S1 to S8
References

18 March 2008; accepted 11 June 2008
10.1126/science.1157996

Measurement of the Distribution of Site Enhancements in Surface-Enhanced Raman Scattering

Ying Fang,* Nak-Hyun Seong,* Dana D. Dlott†

On nanotextured noble-metal surfaces, surface-enhanced Raman scattering (SERS) is observed, where Raman scattering is enhanced by a factor, $\overline{G}$, that is frequently about one million, but underlying the factor $\overline{G}$ is a broad distribution of local enhancement factors, η. We have measured this distribution for benzenethiolate molecules on a 330-nanometer silver-coated nanosphere lattice using incident light of wavelength 532 nanometers. A series of laser pulses with increasing electric fields burned away molecules at sites with progressively decreasing electromagnetic enhancement factors. The enhancement distribution $P(\eta)d\eta$ was found to be a power law proportional to $(\eta)^{-1.75}$, with minimum and maximum values of 2.8×10^4 and 4.1×10^{10}, respectively. The hottest sites ($\eta > 10^9$) account for just 63 in 1,000,000 of the total but contribute 24% to the overall SERS intensity.

Molecules on nanotextured noble-metal surfaces (*1*, *2*) or nanoparticle aggregates (*3*, *4*) frequently evidence giant Raman scattering cross sections. This surface-enhanced Raman scattering (SERS) effect has enabled a variety of chemical sensing applications (*5*), including the detection of single molecules by Raman scattering (*6*, *7*). The average value of the Raman enhancement, $\overline{G}$, is frequently about 10^6 compared with molecules without a SERS substrate. "Hot" spots where the local field enhancement η is 10^9 or more have been detected by searching nanoparticle aggregates with powerful microscopes (*6*–*8*). At a hot spot it is possible to measure the Raman spectrum of single molecules (*6*–*8*). The existence of hot spots suggests that the average enhancement $\overline{G}$ represents a broad distribution of microscopic enhancement factors, so a SERS signal might result from a few molecules at hot sites or the preponderance of molecules at "cold" sites (*9*).

SERS mechanisms may involve electromagnetic enhancement, chemical enhancement, or resonance enhancement. The benzenethiolate (BT) molecule is frequently used as a probe of electromagnetic enhancement. BT forms a densely packed, well-ordered self-assembled monolayer (SAM) on Ag (*10*) of the type frequently used in chemical sensing measurements. Because BT has weak electronic interactions with metal surfaces and does not absorb at the laser wavelength, the chemical and resonance enhancements are unimportant. In electromagnetic enhancement, an incident laser field, E_{in}, excites surface plasmons to create a complex pattern of spatially varying electromagnetic fields (*11*, *12*). At any location, the local field is gE_{in}, where g is the local enhancement factor. For the purposes of this study, it is sufficient to use the approximation (*13*) that

---

School of Chemical Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: dlott@scs.uiuc.edu

the local Raman cross-section enhancement $\eta = \sigma_R^e/\sigma_R^0 = g^4$, where σ_R^e and σ_R^0 are enhanced and unenhanced cross sections, respectively (*12*). The distribution of local field enhancements will be denoted $P(g)dg$, and the distribution of Raman cross-section enhancements is $P(\eta)d\eta = P(g)dg$.

Electromagnetic calculations have been used to study what sorts of metallic nanostructures create hot sites and how large η could conceivably be at the hottest sites (*9*, *14*–*18*). One model system for computational studies consists of two closely spaced metal nanoparticles. In the gap between two closely spaced Ag nanoparticles (*9*, *18*–*20*), η may be as large as 10^{11}. The distribution $P(\eta)d\eta$ has been computed for 25-nm-diameter Ag spheres separated by 2 nm by La Rue and co-workers (*9*). The computed distribution was a power law proportional to $\eta^{-1.135}$, $\overline{G}$ was 6.7×10^7, and the smallest and largest enhancements were about 1.5×10^3 and 2×10^{10}, respectively.

A method to measure $P(\eta)d\eta$ could have several applications. Ordinary Raman measurements are sensitive only to $\overline{G}$, which can lead to difficulties in quantifying SERS signals in analytical applications (*21*, *22*). A great deal of effort has gone into developing methods to fabricate useful SERS materials (*23*), and a knowledge of $P(\eta)d\eta$ would be more useful than a measurement of $\overline{G}$ to optimize a fabrication technology (*22*). However, it is not possible to determine $P(g)dg$ or $P(\eta)d\eta$ distributions by using linear Raman ensemble measurements. This task requires either a nonlinear spectroscopy or a molecule-by-molecule census. But single-molecule measurements would not be sensitive enough to observe the colder sites. In studies of molecules in disordered media, photochemical hole burning (PHB), a nonlinear method that involves a burning pulse and a probing pulse (*24*), has been an effective method for extracting distributions of molecular energy level spacings (*25*), so an analogous method might be used to measure $P(\eta)d\eta$. Photobleaching of adsorbed dye molecules (*9*, *26*) has been suggested as a method of determining $P(\eta)d\eta$ because the photobleaching efficiency increases in regions of greater local field enhancement. Vibrational pumping of adsorbed dyes has also been suggested in this context (*27*).

Our method for determining $P(\eta)d\eta$ uses the electric fields from powerful nonresonant laser pulses to photochemically damage molecules adsorbed on a SERS substrate. This type of photodamage is characterized by a sharp electric field threshold, E_{th}. The value of E_{th} depends on the type of molecule, the wavelength, and the pulse duration, but it is typically 1 to 10 GV m^{-1} (*28*). This sharp threshold behavior, which photobleaching does not exhibit, greatly simplifies the theoretical analysis needed to extract $P(\eta)d\eta$ from experiment. For a given E_{in}, photodamage will occur only at sites where $gE_{\mathrm{in}} \geq E_{\mathrm{th}}$, so as the PHB laser field is increased molecules with the largest g burn away first, followed by molecules at sites with progressively smaller g. Meanwhile, the sample loses Raman intensity at a rate proportional to $\eta = g^4$ (*26*). The disappearance of the photodamaged molecules and the appearance of photoproducts can be monitored via Raman spectroscopy with a weak probe laser. In this study, the PHB method is illustrated by using a SERS substrate composed of an Ag film on nanospheres (AgFON) (*3*–*5*, *23*, *29*) having an adsorbed layer of BT.

The experimental concept is depicted in Fig. 1 (*30*). A BT SAM was deposited on AgFON (330-nm nanospheres coated by 150-nm Ag) fabricated with use of methods developed by the van Duyne group (*23*, *29*). The 1-ps duration, 532-nm PHB laser pulses were focused to 140 μm (Gaussian $1/e^2$ diameter) spots. The pulse energies ranged from 1 to 1000 nJ, and E_{in} ranged from 10 to 300 MV m^{-1}. A 0.5-mW continuous-wave 532 Raman laser

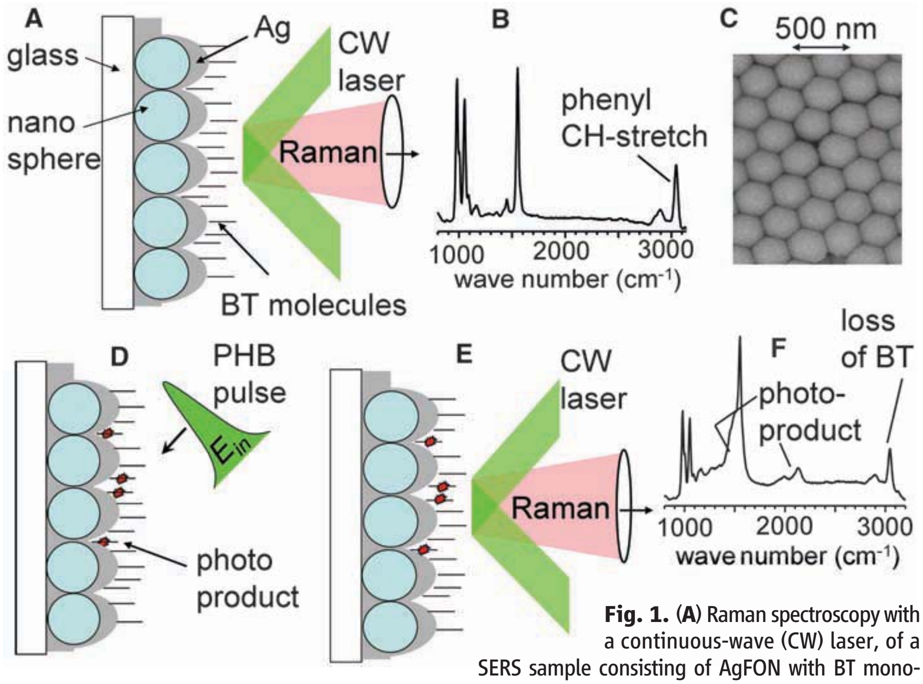


Fig. 1. (**A**) Raman spectroscopy with a continuous-wave (CW) laser, of a SERS sample consisting of AgFON with BT monolayer. (**B**) SERS spectrum of BT. (**C**) Scanning electron micrograph of AgFON surface. (**D**) The sample was exposed to an intense PHB pulse with laser field E_{in}. BT molecules at sites with local field enhancement g were damaged if $gE_{\mathrm{in}} \geq E_{\mathrm{th}}$, where E_{th} is the threshold field needed to damage BT. (**E** and **F**) The Raman spectrum after PHB shows loss of BT plus new transitions from photoproduct molecules. The loss of BT is quantified by using the integrated area of the phenyl CH-stretch transition at 3050 cm^{-1}.

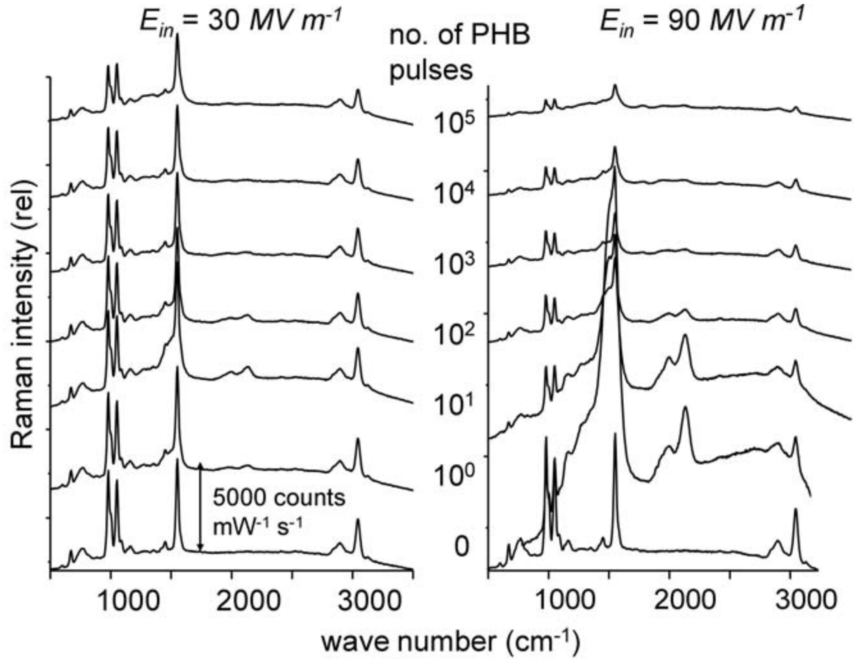


Fig. 2. Raman spectra of BT on an AgFON substrate after the indicated number of PHB pulses, at a lower and higher value of the incident field, E_{in}.

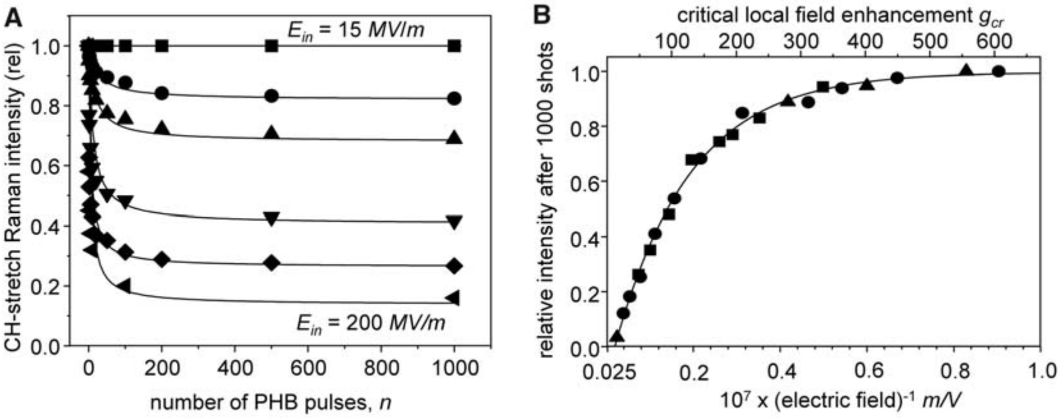


Fig. 3. (**A**) PHB curves for CH-stretch transition of BT molecules on AgFON SERS substrates. From top to bottom, E_{in} = 15, 30, 43, 88, 140, and 200 MV m^{-1}. The smooth curves are calculated from the local field enhancement distribution $P(g)dg$ using the statistical model for PHB of Eq. 1. rel. indicates relative. (**B**) Fraction of BT CH-stretch SERS intensity after 1000 PHB pulses. The different symbols represent measurements on different samples. The PHB curve fits an exponential function (solid line). At given E_{in}, only molecules where the local field enhancement exceeds a critical value g_{cr} (upper abscissa) are damaged by PHB pulses. One-half of the overall SERS signal comes from sites with $g > 100$.

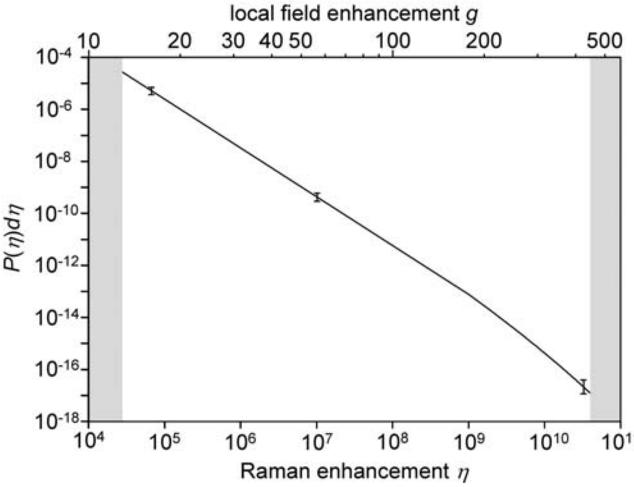


Fig. 4. Measured distribution of SERS enhancement factors $P(\eta)d\eta$ at 532 nm for BT monolayer on AgFON substrate with 330-nm-diameter spheres. The error bars represent estimated errors due to uncertainty in the determination of the ensemble-averaged SERS enhancement $\overline{G}$. Shaded regions denote $\eta_{min} = 2.8 \times 10^4$ and $\eta_{max} = 4.1 \times 10^{10}$.

probed molecules at the center of the PHB beam where E_{in} was spatially uniform to alleviate complications caused by averaging over the spatial beam profile (*26*).

Representative Raman spectra of BT on AgFON after exposure to PHB pulses are shown in Fig. 2. Before PHB the observed spectra agree well with the literature, and the largest peaks give the expected Raman intensity of ~5000 counts mW^{-1} s^{-1} (*3*). The bulk Raman enhancement factor $\overline{G}$ was determined by using the established method (*31*) of comparing the BT SERS intensity to a thin layer of liquid benzenethiol. We used a geometric model for the Ag surface, where the area is a factor $S_{Ag} = 2\pi(R+h)/\sqrt{3R}$ greater than a flat surface (R is the nanosphere radius and h the Ag coating thickness) and the measured value of 3.3×10^{14} cm^{-2} for the molecular packing density on flat Ag(111) (*10*), to obtain $\overline{G} = 8.5 \times 10^5$. Given the possibility of error in our geometric model and the packing density, we assigned a wide error bound of ~50% so that $\overline{G} = 9 \times 10^5 \pm 4 \times 10^5$.

As shown in Fig. 2, in the range of 10 to 100 PHB pulses the BT Raman transitions lose intensity, and new photoproduct transitions grow in, most prominently on the red edge of the 1550 cm^{-1} CH bend and in the 2000 to 2200 cm^{-1} range of carbonyl stretching. On the basis of studies in an O_2-depleted atmosphere, we believe that the observed photoproduct results from field ionization of BT to produce a transient species that later reacts with ambient O_2. Raman signals from the photoproduct have about the same SERS enhancements as BT. The dramatic spectrum observed after a single PHB pulse at $E_{in} = 9 \times 10^7$ V m^{-1} indicates large SERS enhancement of the photoproduct. This result is compelling evidence that PHB pulses damage BT but not the SERS substrate. As support for this conclusion, we observed that after 10^3 intense PHB pulses with E_{in} = 200 MV m^{-1} there was no evident effect on the AgFON surface plasmon resonance.

As the PHB pulse irradiation continued, after 10^3 to 10^4 pulses the photoproduct transitions disappeared. This continued photodamage of the photoproduct further emphasizes that photoproduct molecules reside in local environments that continue to possess large local field enhancements. In the 10^3 to 10^5 pulse regime, BT molecules that were not photodamaged continued to evidence the same Raman spectrum as unirradiated molecules. With subsequent irradiation, after 10^5 to 10^6 high-field pulses, the BT Raman spectral lineshapes evolved and broadened. We believe that this spectral evolution is associated with gradual laser substrate damage.

To quantify the extent of photodamage, we focused on BT transitions well separated from the photoproduct, specifically the integrated area of the 3050 cm^{-1} peak arising from aromatic CH-stretch transitions. The aromatic phenyl moiety of BT is most susceptible to high-field damage, and nonaromatic photoproducts will have their CH-stretch transitions shifted to the 2850 to 2950 cm^{-1} range. Figure 3A shows a series of burning curves based on monitoring the 3050 cm^{-1} CH stretch. The burning curves were not appreciably different when other BT Raman transitions were monitored. The experimental observable in Fig. 3A is the fractional change $I(n)/I_0$ in Raman intensity after n PHB shots at field E_{in}. Although $I(n)$ is quite dependent on n during the first 100 pulses, the intensity drop caused by photodamage is complete at 1000 pulses. To model this behavior, we assumed that each PHB pulse had a constant probability for inducing photodamage. Then, at a site with enhancement factor g irradiated by n pulses at field E_{in}, this probability $P_{PHB}(g;E_{in},n)$ is

$$P_{PHB}(n;g,E_{in}) = 0,\ gE_{in} < E_{th}$$
$$P_{PHB}(n;g,E_{in}) = 1 - \exp[-n/n_0],\ gE_{in} \geq E_{th} \tag{1}$$

where n_0 is the number of pulses needed to damage 1/*e* of the BT molecules. Then $I(n)/I_0$ can be written in terms of the microscopic distribution function $P(g)dg$

$$\frac{I(n)}{I_0} = \frac{\int_{g_{min}}^{g_{max}} P(g)g^4[1 - P_{PHB}(n;g,E_{in})]dg}{\int_{g_{min}}^{g_{max}} P(g)g^4 dg} = \frac{\int_{g_{min}}^{g_{max}} P(g)g^4[1 - P_{PHB}(n;g,E_{in})]dg}{\overline{G}} \tag{2}$$

Table 1. Contribution of the various site enhancements at 532 nm to the overall SERS signal.

Raman enhancement factor η	Percentage of molecules	Percentage contribution to overall SERS signal
$<2.8 \times 10^4$	0	0
2.8×10^4 to 1×10^5	61%	4%
10^5 to 10^6	33%	11%
10^6 to 10^7	5.1%	16%
10^7 to 10^8	0.7%	22%
10^8 to 10^9	0.08%	23%
10^9 to 10^{10}	0.006%	17%
$>10^{10}$	0.0003%	7%

Although it is possible to determine $P(g)dg$ from the data in Fig. 3A, a more accurate determination was made by averaging the results from five fresh regions of the SERS sample, each exposed to 1000 PHB pulses. We can define a critical value of the local field enhancement $g_{cr} = E_{th}/E_{in}$ such that the probability of photodamage is unity at sites with $g \geq g_{cr}$ and zero at sites with $g < g_{cr}$ Then Eq. 2 becomes

$$\frac{I}{I_0} = \frac{\int_{g_{\min}}^{g_{cr}} P(g) g^4 dg}{\overline{G}} = \frac{\int_{\eta_{\min}}^{\eta_{cr}} \eta P(\eta) d\eta}{\overline{G}} \qquad (3)$$

Equation 3 shows that I/I_0 is a function of $g_{cr} \propto 1/E_{in}$, so in Fig. 3B we plotted I/I_0 versus $1/E_{in}$. To determine $P(g)dg$ from Fig. 3B, we could take the numerical derivative of the data and divide by g^4. However the data in Fig. 3B were unexpectedly well fit by an exponential function, leading to an empirical analytical form for $P(g)dg$ and $P(\eta)d\eta$,

$$P(\eta)d\eta = \frac{A}{\eta^{1.75}} \exp\left(\frac{-\eta^{0.25}}{A'}\right) \qquad (4)$$

The value of the constant A in Eq. 4 is determined by the constraint $\int P(g)g^4 dg = \overline{G} = 9 \times 10^5 \pm 4 \times 10^5$. The value of the constant A′ in Eq. 4 is determined by varying the value of E_{th} subject to the normalization condition $\int P(\eta)d\eta = 1$; the best fit was obtained with $E_{th} = 6.7 \pm 0.6$ GV m^{-1}. Knowing E_{th} for BT, with the relation $E_{in} g_{cr} = E_{th}$, we obtain the useful upper abscissa in Fig. 3B. Lastly, we can fit the burning curves in Fig. 3A by using Eqs. 1 and 2 with one additional parameter $n_0 = 10$, which characterizes the number of PHB pulses needed to complete the photodamage process.

In order to properly normalize $P(\eta)d\eta$, we need to know the minimum and maximum enhancement values, η_{max} and η_{min}. To determine η_{max}, we need to determine the weakest pulse that causes the first detectable photodamage. From Fig. 3B, this E_{in} value corresponds to $g_{cr} = 450$ or $\eta_{max} = 4.1 \times 10^{10}$. To determine η_{min}, we need to determine the weakest E_{in} that photodamages every molecule on the surface. Unfortunately, we could not do this without damaging the SERS substrate, so we could not extend Fig. 3A below $g_{cr} = 20$ and $\eta_{min} = 1.6 \times 10^5$. It seems reasonable to extrapolate the exponential in Fig. 3B to where it intersects the abscissa at $g_{cr} = 13$ and $\eta_{min} = 2.8 \times 10^4$. The resulting $P(\eta)d\eta$ at 532 nm is shown in Fig. 4. The distribution of site enhancements is a power law with $\eta^{-1.75}$ dependence, but at the largest enhancements, $\eta > 10^9$, the distribution drops off even more steeply. Thus, our measured distribution for a periodic lattice of 330-nm adjacent nanospheres falls off faster than the $\eta^{-1.135}$ dependence obtained theoretically by La Rue and co-workers (*9*). The La Rue calculation (*9*) refers to a pair of 25-nm spheres separated by 2 nm rather than a lattice of spheres, but interestingly the authors speculated that for a collection of spheres the main difference would be a faster drop off at large η such as we see in Fig. 4. The fraction of molecules at each enhancement and the contributions of the different site enhancement factors η to the overall SERS signal are given in Table 1, derived by using Fig. 3B and numerical integration of Fig. 4. The distribution we present applies to AgFON substrates with the specific geometry used here and might be quite different for other SERS materials.

This study used 1-ps, 532-nm PHB pulses, but we know no reason why other PHB pulse durations and wavelengths could not be used. Ultrashort pulses may not be required but are desirable because at a given E_{in} ultrashort pulses pose less risk of substrate damage. We believe that the largest source of experimental error is our error in determining $\overline{G}$. The resulting uncertainty in $P(\eta)d\eta$ is illustrated by the error bars in Fig. 4. Our measured distribution terminates at $\eta_{max} = 4.1 \times 10^{10}$. There may be a small number of hotter sites, but they represent such a tiny fraction of the overall SERS intensity that if they existed we could not detect them. The value $\eta_{min} = 2.8 \times 10^4$ is based on extrapolation, and if there were more or fewer cold sites than the extrapolation indicates, the fraction of hot sites in Table 1 would become proportionately smaller or larger.

Table 1 answers many questions about the inhomogeneous nature of the AgFON SERS substrate. The coldest sites ($\eta < 10^5$) contain 61% of the molecules but contribute just 4% of the overall SERS intensity. The hottest sites ($\eta > 10^9$) comprise just 63 molecules per million but contribute 24% of the overall SERS intensity.

Our PHB technique should be capable of measuring distributions on other SERS materials provided these substrates can withstand PHB pulses. A particularly interesting system involves stripping away the Ag-coated nanospheres (*23*, *31*), leaving behind a periodic array of nanotriangles. The reduced area of the nanotriangle substrate, about 3% of the AgFON surface area for BT binding, reduces the overall SERS intensity but increases the average enhancement $\overline{G}$ of the remaining surface. If we assume the stripping process removes the coldest 97% of the BT molecules, then from the $P(\eta)d\eta$ distribution found here we would predict the average enhancement for BT on nanotriangles to be $\overline{G} = 2 \times 10^7$, about 20 times greater than for AgFON. Although we cannot directly compare the nanotriangle experiments to the present work because the Ag thickness, sphere diameters, and wavelengths were slightly different, with a substrate that performed optimally at 625-nm laser wavelength McFarland *et al.* (*31*) obtained $\overline{G} = 1.2 \times 10^7$ and with a 670-nm substrate $\overline{G} = 1.4 \times 10^7$.

References and Notes

1. M. Fleischmann, P. J. Hendra, A. J. McQuillan, *Chem. Phys. Lett.* **26**, 163 (1974).
2. D. L. Jeanmaire, R. P. Van Duyne, *J. Electroanal. Chem.* **84**, 1 (1977).
3. C. L. Haynes, R. P. Van Duyne, *J. Phys. Chem. B* **105**, 5599 (2001).
4. K. A. Willets, R. P. Van Duyne, *Annu. Rev. Phys. Chem.* **58**, 267 (2007).
5. C. R. Yonzon, X. Y. Zhang, J. Zhao, R. P. Van Duyne, *Spectroscopy* **22**, 42 (2007).
6. K. Kneipp, H. Kneipp, I. Itzkan, R. R. Dasari, M. S. Feld, *Chem. Phys.* **247**, 155 (1999).
7. J. T. Krug II, G. D. Wang, S. R. Emory, S. Nie, *J. Am. Chem. Soc.* **121**, 9208 (1999).
8. J. A. Dieringer, R. B. Lettan, K. A. Scheidt, R. P. van Duyne, *J. Am. Chem. Soc.* **129**, 16249 (2007).
9. E. C. Le Ru, P. G. Etchegoin, M. Meyer, *J. Chem. Phys.* **125**, 204701 (2006).
10. J. Y. Gui *et al.*, *Langmuir* **7**, 955 (1991).
11. M. I. Stockman, *Phys. Rev. E* **56**, 6496 (1997).
12. M. Moskovits, in *Surface-Enhanced Raman Scattering: Physics and Applications*, vol. 103 of *Topics in Applied Physics*, K. Kneipp, M. Moskovits, H. Kneipp, Eds. (Springer-Verlag, Berlin, 2006), pp. 10–18.
13. E. C. Le Ru, P. G. Etchegoin, *Chem. Phys. Lett.* **423**, 63 (2006).
14. P. K. Aravind, H. Metiu, *J. Phys. Chem.* **86**, 5076 (1982).
15. P. K. Aravind, H. Metiu, *Surf. Sci.* **124**, 506 (1983).
16. P. K. Aravind, A. Nitzan, H. Metiu, *Surf. Sci.* **110**, 189 (1981).
17. N. Liver, A. Nitzan, J. I. Gersten, *Chem. Phys. Lett.* **111**, 449 (1984).
18. G. C. Schatz, M. A. Young, R. P. van Duyne, in *Surface-Enhanced Raman Scattering: Physics and Applications*, vol. 103 of *Topics in Applied Physics*, K. Kneipp, M. Moskovits, H. Kneipp, Eds. (Springer-Verlag, Berlin, 2006), pp. 19–46.
19. H. X. Xu, J. Aizpurua, M. Käll, *Phys. Rev. E* **62**, 4318 (2000).
20. H. X. Xu, M. Kä ll, *ChemPhysChem* **4**, 1001 (2003).
21. E. C. Le Ru, E. Blackie, M. Meyer, P. G. Etchegoin, *J. Phys. Chem. C* **111**, 13794 (2007).
22. M. Natan, *Faraday Discuss.* **132**, 321 (2006).
23. X. Y. Zhang, A. V. Whitney, J. Zhao, E. M. Hicks, R. P. Van Duyne, *J. Nanosci. Nanotech.* **6**, 1920 (2006).
24. S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford Univ. Press, New York, 1995).
25. W. W. Parson, *Modern Optical Spectroscopy* (Springer, Berlin, 2007).
26. B. Pettinger, B. Ren, G. Picardi, R. Schuster, G. Ertl, *J. Raman Spectrosc.* **36**, 541 (2005).
27. P. G. Etchegoin, E. C. Le Ru, R. C. Maher, L. F. Cohen, *Phys. Chem. Chem. Phys.* **9**, 4923 (2007).
28. K. Toyota, S. Nakashima, T. Okada, *Chem. Phys. Lett.* **323**, 323 (2000).
29. X. Zhang, M. S. Young, M. A. Lyandres, R. P. Van Duyne, *J. Am. Chem. Soc.* **127**, 4484 (2005).

30. Materials and methods are detailed in supporting online material available at *Science* Online.
31. A. D. McFarland, M. A. Young, J. A. Dieringer, R. P. van Duyne, *J. Phys. Chem. B* **109**, 11279 (2005).
32. This material is based on work supported by NSF under award DMR 0504038, the Air Force Office of Scientific Research under award FA9550-06-1-0235, and the Army Research Office under award W911NF-05-1-0345. Electron microscopy was carried out in the Center for Microanalysis of Materials, University of Illinois, which is supported by the U.S. Department of Energy under grant DEFG02-91ER45439.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1159499/DC1
Materials and Methods
Fig. S1
References

23 April 2008; accepted 16 June 2008
Published online 26 June 2008;
10.1126/science.1159499
Include this information when citing this paper.

Patagonian Glacier Response During the Late Glacial–Holocene Transition

Robert P. Ackert Jr.,[1]* Richard A. Becker,[2] Brad S. Singer,[2] Mark D. Kurz,[3] Marc W. Caffee,[4] David M. Mickelson[2]

Whether cooling occurred in the Southern Hemisphere during the Younger Dryas (YD) is key to understanding mechanisms of millennial climate change. Although Southern Hemisphere records do not reveal a distinct climate reversal during the late glacial period, many mountain glaciers readvanced. We show that the Puerto Bandera moraine (50°S), which records a readvance of the Southern Patagonian Icefield (SPI), formed at, or shortly after, the end of the YD. The exposure age (10.8 ± 0.5 thousand years ago) is contemporaneous with the highest shoreline of Lago Cardiel (49°S), which records peak precipitation east of the Andes since 13 thousand years ago. Absent similar moraines west of the Andes, these data indicate an SPI response to increased amounts of easterly-sourced precipitation—reflecting changes in the Southern Westerly circulation—rather than regional cooling.

First detected in pollen records from northern Europe, the Younger Dryas (YD) chronozone is perhaps best expressed in Greenland ice cores (*1*) that show an abrupt return to near-glacial conditions in the North Atlantic region during the last deglaciation between 12.9 and 11.6 thousand years ago (ka). Less extreme, but extensive, cooling occurred throughout the Northern Hemisphere (*2*). Comparison of temperature records from Greenland and Antarctic ice cores during the last deglaciation indicate antiphased behavior, however (*3*); the Antarctic Cold Reversal (ACR) preceded the YD interval and Antarctic warming occurred during the YD, peaking at ~11 ka. Determining whether the Southern Hemisphere mid-latitudes follow a North Atlantic or an Antarctic signal remains key to deciphering the mechanisms responsible for millennial and abrupt climate change (*4*).

In general, climate records from New Zealand and Patagonia lack evidence of a distinct climate reversal during the late glacial period [see the supporting online material (SOM) text]. Yet there is clear evidence for prominent glacial readvances in both New Zealand's Southern Alps (*5*, *6*) and the Southern Andes after ~15 ka (*7*). In New Zealand, the Waiho Loop moraine, which was previously inferred to be of YD age on the basis of ^{14}C ages of reworked wood, may be significantly younger in accordance with cosmogenic surface-exposure ages of moraine boulders (*8*, *9*). ^{14}C ages provide only broadly limiting constraints on late glacial advances in Patagonia. Thus, both the timing and synchrony of Southern Hemisphere glacial readvances remain poorly constrained.

The Puerto Bandera moraines at Lago Argentino, Argentina (50°S), consist of two terminal moraine belts (PB I and PB II) that can be distinguished on the basis of field observations and interpretation of remote imagery (*10*). The moraines occur along both the northern and southern shore of the lake with the younger (PB II) moraine crosscutting the older (PB I) near the present shoreline (Fig. 1C). Lying within the Patagonian steppe, where precipitation is currently only ~200 mm/year, the moraines record readvances of Southern Patagonian Icefield (SPI) outlet glaciers that extended 40 km beyond the nearest present-day ice margins.

Three ^{14}C dates of peat overlying gravel from an abandoned outlet channel in the nearby Brazo Rico moraine (Fig. 1C) with a weighted mean age of 11.7 ± 0.3 ka (*10*–*12*) have been taken as a minimum age of the Puerto Bandera moraines (*13*). Although the ages are consistent with a YD age for the Brazo Rico moraine, Mercer considered this interpretation unlikely because it requires a very rapid retreat of the ice front at the end of the YD (*11*); peat could not begin to accumulate in the outlet channel until the glacier had retreated to the position of the present Perito Moreno Glacier, thereby opening a lower outlet to Lago Argentino along its terminus (Fig. 1C). Strelin and Malagnino report ^{14}C ages of reworked peat in lateral moraines in Brazo Norte (Fig. 1C) (*10*). An age of 15.5 ± 2.4 ka was obtained from a PB I moraine and an age of 12.9 ± 2.1 ka was obtained from a PB II moraine. The ^{14}C ages of reworked peat show that the SPI had retreated within the Cordillera at these times and indicate that the terminal Puerto Bandera moraines record a large-scale readvance of the SPI.

To better constrain the age of the Puerto Bandera moraines and to evaluate Southern Hemisphere climate during the YD interval, we determined the exposure ages of 18 moraine boulders using cosmogenic ^{36}Cl and ^{10}Be. Samples were collected from the outermost moraine crests on both the northern and southern side of the lake. Fifteen samples are from PB I and three samples are from PB II (Fig. 1C). Boulders were selected on the basis of size, shape, quartz content (for the ^{10}Be samples), and lack of weathering in order to reduce geologic uncertainties. The ^{36}Cl and ^{10}Be production rates have been scaled for elevation, latitude, and atmospheric pressure (*14*). Adjustments for erosion and secular variation of the magnetic field were unnecessary. We estimate that uncertainties in our exposure ages, including scaling factors, are ~6% (*15*).

The arithmetic means of the ^{10}Be and ^{36}Cl exposure ages are 10.8 ± 2.1 ka and 10.8 ± 2.9 ka, respectively (Table 1). More involved statistical treatment relies on the assumption that the exposure ages comprise a normal distribution, specifically, that all samples have an identical simple exposure history and that scatter in the ages results from normally distributed analytical uncertainties. We tested this assumption using a χ^2 analysis. We found the youngest ^{36}Cl and ^{10}Be exposure ages to be outliers based on the low probability that the samples come from a single normal distribution (*16*). The weighted mean of the remaining ^{10}Be ages is 10.8 ± 0.6 ka; that of the remaining ^{36}Cl ages is 10.9 ± 0.9 ka (Table 1 and tables S2 and S3). All 16 samples yielded an inverse-variance weighted mean age of 10.8 ± 0.5 ka (the mean square weighted deviation is 0.80). Given that the scatter in the ages can be attributed mostly to normally distributed analytical errors, we inferred that

[1] Department of Earth and Planetary Science, Harvard University, 20 Oxford Street, Cambridge, MA 02138, USA. [2] Department of Geology and Geophysics, University of Wisconsin–Madison, 1215 West Dayton Street, Madison, WI 53706, USA. [3] Woods Hole Oceanographic Institution, Clark 419, MS #25, Woods Hole, MA 02543, USA. [4] Purdue Rare Isotope Measurement Laboratory, Purdue University, 525 Northwestern Avenue, West Lafayette, IN 47907–2036, USA.

*To whom correspondence should be addressed. E-mail: rackert@fas.harvard.edu

the moraines stabilized within several hundred years (*15*).

We interpreted the exposure age of the PB I moraine as the end of deposition and the start of ice recession and, thus, a fundamental change in the regional climate. The magnitude of the climate change represented by the PB I glacial advance is indicated by the equilibrium-line altitude (ELA) depression in the Rio Mitre Valley (Fig. 1C) adjacent to the PB moraines. Two ^{36}Cl ages on an alpine glacier moraine at 240 m elevation indicate that the glacier advanced contemporaneously with the PB I. There are no glaciers within the drainage at present; we estimated that the ELA is 1450 ± 50 m based on Shuttle Radar Topography Mission and Landsat 7 data of a glacier in the pre-Cordillera 40 km north of Lago Argentino. Using an accumulation area ratio of 0.65 ± 0.05, we found that the paleo-ELA was 500 to 600 m lower (*15*).

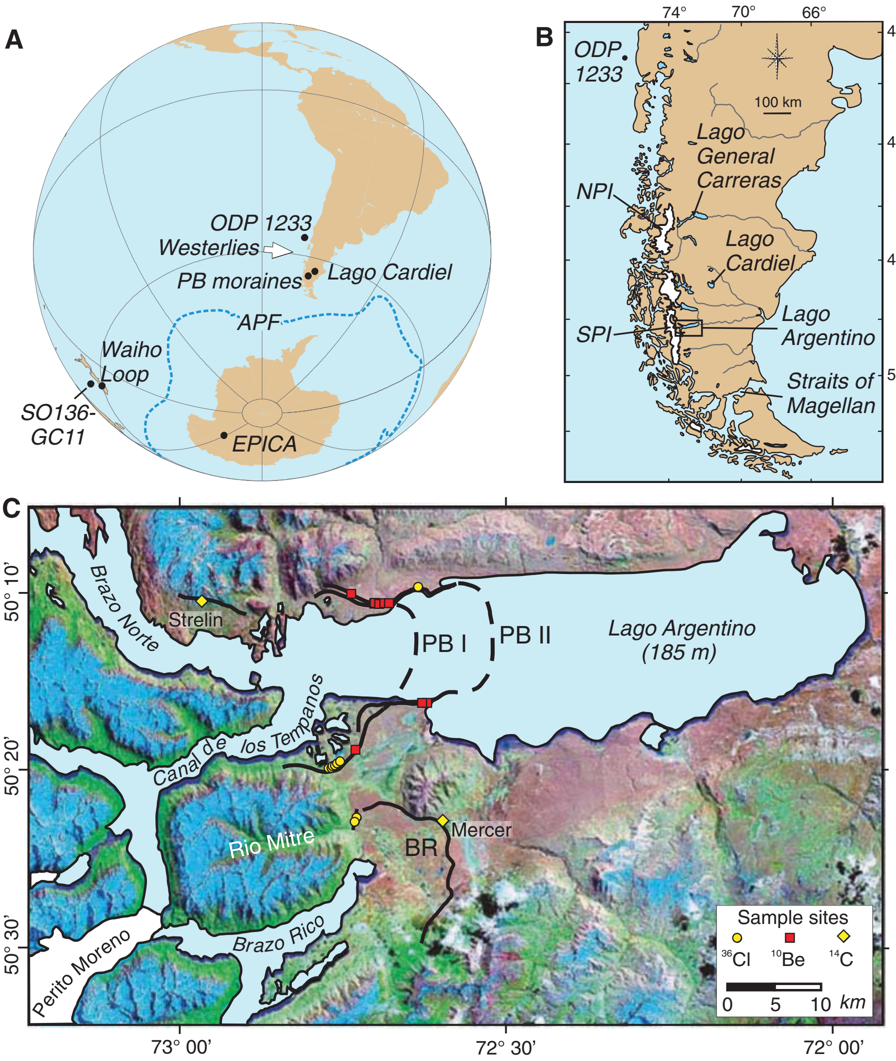


Fig. 1. (**A**) Location of key sites in the Southern Hemisphere. (**B**) Location map of Patagonia showing the NPI, SPI, and key climate records. (**C**) Landsat image of the Lago Argentino area annotated to show the Puerto Bandera moraine crests (dashed and solid lines labeled PB), Brazo Rico moraine (solid line labeled BR), and locations of dated samples.

Table 1. Exposure ages of Puerto Banderas Moraine boulders assuming an erosion rate of zero. Ages are in thousands of years with ±2σ analytical uncertainties.

Sample	^{10}Be age	Sample	^{36}Cl age
PBS-04-05	11.3 ± 1.6	PAT-98-001	13.4 ± 3.4
PBN-04-15	10.5 ± 1.6	PAT-98-002	11.7 ± 2.0
PBS-04-19	11.3 ± 2.6	PAT-98-005	12.1 ± 4.6
PBS-04-30	9.8 ± 1.4	PAT-98-050	9.2 ± 3.8
PBN-04-60	10.5 ± 1.3	PAT-98-051	9.6 ± 1.8
PBN-04-61	11.0 ± 2.6	PAT-98-054	10.7 ± 2.4
PBN-04-62	8.5 ± 2.1	PAT-98-057	8.5 ± 1.4
PBN-04-66	12.2 ± 2.2	PAT-98-100	10.8 ± 2.0
PBN-04-67	11.9 ± 2.2	PAT-98-101	11.6 ± 2.8
Arithmetic mean	10.8 ± 2.1		10.8 ± 2.9
Weighted mean of 8 oldest ages	10.8 ± 0.6		10.9 ± 0.9
Weighted mean of 16 oldest ages			10.8 ± 0.5

The exposure age of 10.8 ± 0.5 ka is younger than the YD, as expressed in Greenland ice (Fig. 2A). Adopting the more conservative estimate of scaling uncertainty that is typically assumed (±10%), the exposure age would fall within the end of the YD; this would not alter our conclusions. In any case, our results are inconsistent with previous interpretations that the PB I moraine is older than the YD, because of limiting ^{14}C dates of neighboring moraines (*10*, *17*). Considering the limiting ^{14}C age of 12.9 ± 2.1 ka, the duration of the PB I readvance was only 1000 to 2000 years. As shown below, this period is characterized in regional climate records by its increasing precipitation and temperature, which culminated ~11 ka.

The exposure age of the PB I moraine boulders coincides with the ^{14}C age of the highest shoreline at Lago Cardiel (49°S), a closed lake basin east of the Andes (Fig. 1, A and B). Lake levels rose from near desiccation (–73 m below present) after 13.1 ± 0.2 ka (*18*) and peaked at levels 55 m above the present shoreline about 10.8 ± 0.3 ka (*19*, *20*) (Fig. 2F). The desiccation event, which occurred near the end of the ACR, indicates extreme aridity east of the Andes during that time. The high lake levels, which are not seen either during the late glacial period or after the middle Holocene, indicate that precipitation east of the Andes peaked at around 11 ka. West of the Andes, but north of the study area (41°S), an alkenone-based sea surface temperature (SST) record at Ocean Drilling Program (ODP) site 1233, from a high-sedimentation-rate core on the continental slope (*21*) (Fig. 1, A and B, and Fig. 2C), closely resembles the temperature records from Antarctic ice cores (Figs. 1A and 2G). Temperature rose through the YD and reached a maximum by 12 ka. Salinity rose through the ACR and fell through the YD with minimum values coinciding with the maximum lake levels at Lago Cardiel (Fig. 2, E and F). The salinity decrease, as well as the high Fe concentrations (Fig. 2D), is consistent with increased rainfall, erosion, and runoff into fjords west of the southern Andes and the Patagonian ice fields.

Pollen and charcoal records from Patagonia between 38°S and 55°S suggest different climate histories for different latitude bands (*20*). Before 11 ka, maximum precipitation occurred to the north, whereas the middle and southern latitudes were quite dry and somewhat cooler. Between 11 and 10 ka, the north dried out, whereas maximum moisture was focused at the mid-latitudes, though the south remained dry. This pattern reflects latitudinal migration of the Southern Westerlies and their associated storm tracks (*20*). Thus, regional climate proxies suggest that between 45°S and 50°S temperature warmed and precipitation east of the Andes increased throughout the YD, reaching maximum values by 11 ka.

There is evidence for correlative glacier advances at ~11 ka at other locations east of the Andes that indicates that the PB I moraine is not the result of a glacial surge unrelated to climate. Two boulders with exposure ages of 10.5 ± 2.6 ka and 12.2 ± 3.0 ka on a delta 100 m above Lago General Carreras (46°S) are consistent with diversion of the outlet (the Rio Baker) by glaciers advancing eastward from the Northern Patagonian Icefield (NPI) (*22*). In the Rio Bayo valley, draining the northern NPI, exposure ages of boulders resting on striated bedrock are 11.4 ± 1.8 ka and 10.5 ± 1.6 ka (*23*). In the Cordilleran foothills north of Lago Argentino, minimum ^{14}C ages of alpine glacier moraines in the Rio Guanaco drainage are provided by rootlets in a channel (10.7 ± 0.2 ka) and peat in a kettle (10.9 ± 0.2 ka) (*7*). In contrast, on the west side of the Andes at these latitudes, there is neither evidence of a glacial advance (*24*) nor of climate change in other proxies (*25*).

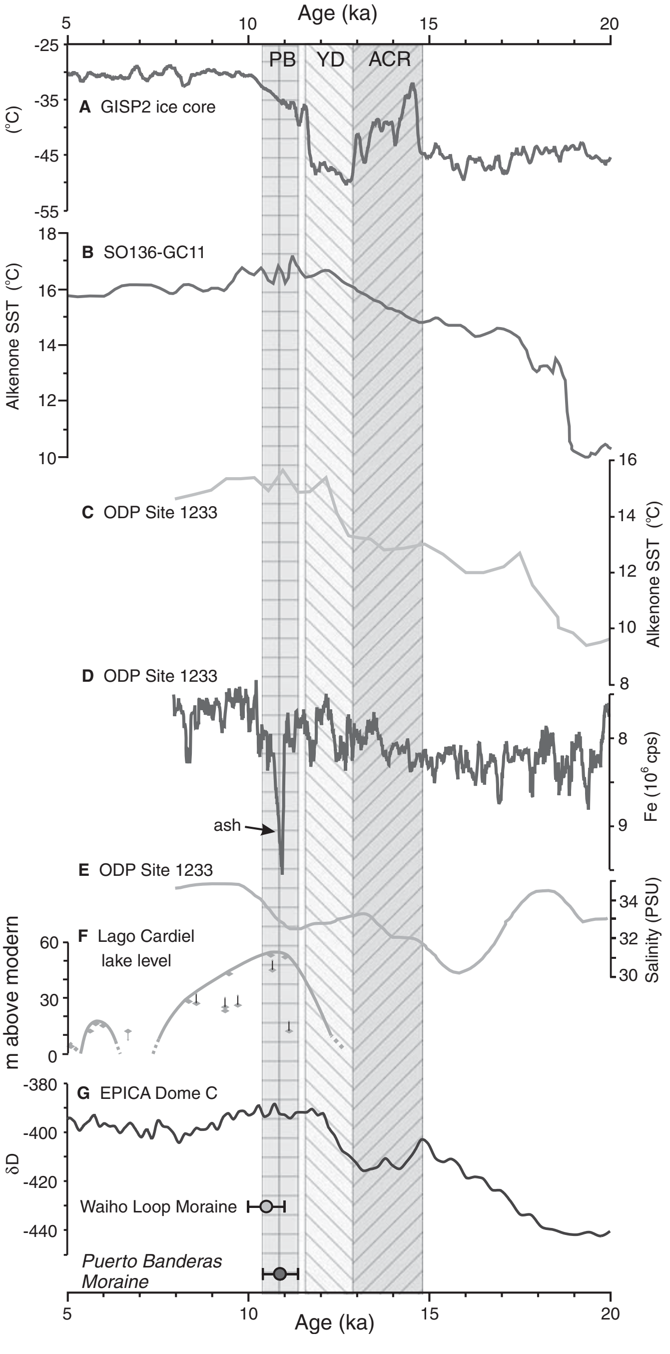


Fig. 2. Regional climate records compared with mean exposure age of PB I boulders, organized north to south. (**A**) δ^{18}O-based temperature of GISP2 (Greenland) ice core (*1*). (**B**) SST at SO136-GC11 off New Zealand (*8*). (**C** to **E**) SST, Fe content, and salinity at ODP Site 1233 (*21*). An Fe-rich volcanic ash bed was deposited during a broader increase in Fe from 11.5 to 10.0 ka. (**F**) Lago Cardiel (49°S) shoreline elevation (*19*). (**G**) δD, a measure of the isotopic fractionation of hydrogen in water (ice); more negative δD values correspond to lower temperatures at the site of the European Project for Ice Coring in Antarctica (EPICA) Dome C (Antarctic) ice core (*26*).

The surface exposure age of 10.8 ± 0.5 ka is younger than the mean ^{14}C age of 11.7 ± 0.3 ka from the Brazo Rico outlet channel (Fig. 1C), which indicates that peat was accumulating in the channel before deposition of the PB I moraine boulders. Thus, we infer that, rather than being contemporaneous, the Brazo Rico moraine predates the PB moraines (SOM text).

In summary, multiple independent paleoclimate records suggest that the extensive glacier advance that occurred at Lago Argentino 10.8 ± 0.5 ka on the dry, eastern side of the Andes (50°S) was a response to dramatically increased precipitation rather than dramatic cooling, such as in the North Atlantic region. In this sector of Patagonia, the YD was characterized by increasing easterly sourced precipitation that peaked at ~11 ka with an associated ELA depression of ~550 m. The moisture was likely carried by southeasterly circulation around a lee-side low-pressure trough resulting from the focusing of the Southern Westerlies between 45°S and 50°S. The ACR was a period of aridity and cooler climate in which the ice caps retreated within the Cordillera. This contrasts with the glacial chronology at the Straits of Magellan at 52°S (*4*) and Lago Buenos Aires at 46°S (*16*), where glaciers advanced during the ACR.

Combined with similar climate records from New Zealand (*8*) (Fig. 2B), the Patagonian moraine chronology indicates that in the southern mid-latitudes, glaciers were responding to marked increases in precipitation resulting from changes in the position and intensity of the Southern Westerly circulation, rather than hemispheric cooling during the YD. That accumulation, rather than summer ablation (temperature), dominates the mass balance of these glaciers is probably a consequence of the extremely high accumulation rates and low seasonality as compared with glaciers in more continental climate regimes. Thus, although apparently synchronous climate changes occurred in the southern mid-latitudes during the YD, the glacier advances were not due to worldwide cooling. Rather, these climate changes show a correspondence with trends in Antarctic ice cores wherein glacier advances in the Patagonian steppe (50°S) and New Zealand occurred during peak temperatures in Antarctica rather than during the ACR.

References and Notes

1. R. B. Alley, *Quat. Sci. Rev.* **19**, 213 (2000).
2. D. M. Peteet, *Quat. Int.* **28**, 93 (1995).
3. T. Blunier *et al.*, *Nature* **394**, 739 (1998).
4. D. E. Sugden, M. J. Bentley, C. J. Fogwell, N. R. J. Hulton, R. D. McCulloch, *Geogr. Ann.* **87**, 273 (2005).
5. G. H. Denton, C. H. Hendy, *Science* **264**, 1434 (1994).
6. S. Ivy-Ochs, C. Schlucter, P. W. Kubik, G. H. Denton, *Geogr. Ann.* **81**, 313 (1999).
7. G. Wenzens, *Quat. Res.* **51**, 238 (1999).
8. T. T. Barrows, S. J. Lehman, K. L. Fifield, P. De Dekker, *Science* **318**, 86 (2007).
9. P. J. Applegate, T. V. Lowell, R. B. Alley, *Science* **320**, 746d (2008).
10. J. A. Strelin, E. C. Malagnino, *Quat. Res.* **54**, 339 (2000).
11. J. H. Mercer, *Quat. Res.* **6**, 125 (1976).

12. A. C. Ashworth, *Anthropologie* **98**, 3 (1994).
13. All age uncertainties are reported at 2σ. Radiocarbon ages were converted to calendar years following (*27*).
14. J. O. Stone, *J. Geophys. Res.* **105**, 23 (2000).
15. Materials and methods are available as supporting online material on *Science* Online.
16. D. C. Douglass, B. S. Singer, M. R. Kaplan, D. M. Mickelson, M. W. Caffee, *Quat. Geochron.* **1**, 43 (2006).
17. J. Mercer, *Am. J. Sci.* **266**, 91 (1968).
18. A. Gilli *et al.*, *Terra Nova* **13**, 443 (2001).
19. S. Stine, M. Stine, *Nature* **345**, 705 (1990).
20. V. Markgraf *et al.*, *Holocene* **13**, 581 (2003).
21. F. Lamy *et al.*, *Science* **304**, 1959 (2004).
22. K. J. Turner, C. J. Fogwell, R. D. McCulloch, D. E. Sugden, *Geogr. Ann.* **87A**, 363 (2005).
23. N. E. Glasser, S. Harrison, S. Ivy-Ochs, G. A. T. Duller, P. W. Kubik, *Quat. Res.* **65**, 70 (2006).
24. C. J. Heusser, *J. S. Am. Earth Sci.* **15**, 577 (2002).
25. K. D. Bennett, S. G. Haberle, S. H. Lumley, *Science* **290**, 325 (2000).
26. B. Stenni *et al.*, *Earth Planet. Sci. Lett.* **217**, 183 (2004).
27. P. J. Reimer *et al.*, *Radiocarbon* **46**, 1029 (2004).
28. We thank J. Stone for providing laboratory facilities for $^{36}\mathrm{Cl}$ preparation at the University of Washington. This research was supported by NSF grants to the University of Wisconsin–Madison, Woods Hole Oceanographic Institution, and Purdue University.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5887/392/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Tables S1 to S3

3 March 2008; accepted 16 June 2008
10.1126/science.1157215

Electronic Publication and the Narrowing of Science and Scholarship

James A. Evans

Online journals promise to serve more information to more dispersed audiences and are more efficiently searched and recalled. But because they are used differently than print—scientists and scholars tend to search electronically and follow hyperlinks rather than browse or peruse—electronically available journals may portend an ironic change for science. Using a database of 34 million articles, their citations (1945 to 2005), and online availability (1998 to 2005), I show that as more journal issues came online, the articles referenced tended to be more recent, fewer journals and articles were cited, and more of those citations were to fewer journals and articles. The forced browsing of print archives may have stretched scientists and scholars to anchor findings deeply into past and present scholarship. Searching online is more efficient and following hyperlinks quickly puts researchers in touch with prevailing opinion, but this may accelerate consensus and narrow the range of findings and ideas built upon.

Scholarship about "digital libraries" and "information technology" has focused on the superiority of the electronic provision of research. A recent Panel Report from the U.S. President's Information Technology Advisory Committee (PITAC), "Digital Libraries: Universal Access to Human Knowledge," captures the tone: "All citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge…. In this vision, no classroom, group, or person is ever isolated from the world's greatest knowledge resources" (*1*, *2*). This perspective overlooks the nature of the interface between the user and the information (*3*). There has been little discussion of browsing/searching technology or its potential effect on science and scholarship.

Recent research into the practice of library usage measures the use of print and electronic resources with surveys, database access logs, circulation records, and reshelving counts. Despite differences in methodology, researchers agree that print use is declining as electronic use increases (*4*), and that general users prefer online material to print (*5*). These studies are also in general agreement about the three most common practices used by scientists and scholars who publish. First, most experts browse or briefly scan a small number of core journals in print or online to build awareness of current research (*6*). After relevant articles are discovered online, these are often printed and perused in depth on paper (*7*). A second practice is to search by topic in an online article database. In recent years, the percentage of papers read as a result of browsing has dropped and been replaced by the results of online searches, especially for the most productive scientists and scholars (*8*). Finally, subject experts use hyperlinks in online articles to view referenced or related articles (*6*). Disciplinary differences exist. For example, biologists prefer to browse online, whereas medical professionals place a premium on purchasing and browsing in print. In sum, researchers peruse in print, browse in print or online (*9*), and search and follow citations online. These findings follow from the organization and accessibility of print and online papers. Print holdings reside either in a physical "stack" by journal and topic, arranged historically, or in a "recent publications" area. For print journals, the table of contents—its list of titles and authors—serves as the primary index. Online archives allow people to browse within journals, but they also facilitate searching the entire archive of available journals. In online interfaces where searching and browsing are both options (e.g., 3 ProQuest, Ovid, EBSCO, JSTOR, etc.), the searching option (e.g., button) is almost always placed first on the interface because logs demonstrate more frequent usage. When searched as an undifferentiated archive of papers, titles, abstracts, and sometimes the full text can be searched by relevance and by date. Because electronic indexing is richer, experts may still browse in print, but they search online (*10*).

What is the effect of online availability of journal issues? It is possible that by making more research more available, online searching could conceivably broaden the work cited and lead researchers, as a collective, away from the "core" journals of their fields and to dispersed but individually relevant work. I will show, however, that even as deeper journal back issues became available online, scientists and scholars cited more recent articles; even as more total journals became available online, fewer were cited.

Citation data were drawn from Thompson Scientific's *Science*, *Social Science*, and *Arts and Humanities* Citation Indexes, the most complete source of citation data available. Citation Index (CI) data currently include articles and associated citations from the 6000 most highly cited journals in the sciences, social sciences, and humanities going back as far as 1945, for a total of over 50 million articles. The CI flags more than 98% of its journals with from 1 to 3 of a possible 300 content codes, such as "condensed matter physics," "ornithology," and "inorganic and nuclear chemistry." Citation patterns were then linked with data tracking the online availability of journals from Information Today, Inc.'s *Fulltext Sources Online* (FSO).

FSO is the oldest and largest publication about electronic journal availability. Information Today began publishing FSO biannually in 1998, indicating which journals were available in which commercial electronic archives (e.g., Lexis-Nexis, EBSCO, Ovid, etc.) or if they were available freely on their own Web site, and for how many back issues. Merged together by ISSN (International Standard Serial Number), the CI and FSO data allowed me to capture how article online availability changes the use of published knowledge in subsequent research. FSO's source distinction further allows comparison of print access with the different electronic channels through which scientists and scholars obtained articles—whether a privately maintained commercial portal or the open Internet. The combined CI-FSO data set resulted in 26,002,796

Department of Sociology, University of Chicago, 1126 East 59th Street, Chicago, IL 60615, USA. E-mail: jevans@uchicago.edu

articles whose journals came online by 2006 and a distinct 8,090,813 (in addition to the 26 million) that referenced them. Figure 1 shows the speed of the shift toward commercial and free electronic provision of articles, and how deepening backfiles have made more early science readily available in recent years.

Panel regression models were used to explore the relation between online article availability and citation activity—average historical depth of citations, number of distinct articles and journals cited, and Herfindahl concentration of citations to particular articles and journals—over time (details on methods are in the Supporting Online Material). Because studies show substantial variation in reading and research patterns by area, I used fixed-effect specifications to compare journals and subfields only to themselves over time as their online availability shifted. In this way, the pattern of citations to a journal or subfield was compared when available only in print, in print and online through a commercial archive, and online for free.

The first question was whether depth of citation—years between articles and the work they reference—is predicted by the depth of journal issues online—how many years back issues were electronically available during the previous year when scientists presumably drafted them into their papers. For subfields, this was calculated as years from the first journal's availability. These data were collected in publication windows of 20 years, and so only data from 1965—20 years after the beginning of the data set—were used. For the entire data set, citations pointed to articles published an average of 5.6 years previously (table S1). The average number of years journal articles were available online is only 1.85 (the data go back to 1945), but with a standard deviation of 5 years and a maximum of more than 60 years. Analysis was performed by citation year and within journal or subfield. The standard ordinary least squares (OLS) method for linear regression was used in generating all the results to be described.

All regression models contained variables used to account and statistically control for alternative explanations of why citations might refer to more recent articles. A sequence of integers from 1 to 40, corresponding to citation years 1965 through 2005, was included to account for a general trend of increasing citations over time (the estimates for this variable were always positive and statistically significant, $P < 0.001$). Average number of pages and average number of references in citing articles were both included to account for the possibility that citations are more recent because articles are shorter with fewer references and the earliest ones have been disproportionately "censored" by publishers (estimates for pages were positive but not always significant; those for references were always positive and significant, $P < 0.001$: longer articles with more references referred to earlier work). A measure of the average age of title words was also included in the models to account for the possibility that in recent years, research has concerned more recent concepts or recently discovered (or invented) phenomena. This was calculated by taking the age of each title word within the relevant publication window for the analysis (e.g., prior 20 years) and then multiplying it by a weight for each word i in title j equivalent to $\sum_{i=1}^{k}(1+\ln(tf_{ij}))\times\ln\left(\frac{N}{df_i}\right)$ where tf_{ij} equals the frequency of term i in title j and df_i equals the number of articles in a given year that contain term i out of the total number of annual articles N (*11*). This approach highly weights distinguishing title terms (e.g., *buckeyballs*, *microRNA*) and gives lesser weight to broad area terms (e.g., *gene*, *ocean*) and virtually no weight to universal words (e.g., *and*, *the*). Regression coefficients for the title age measures were always positive and significant ($P < 0.0001$), indicating that titles with older terms referenced earlier articles. Each model also contained a constant with a significant negative estimate.

The graphs in Fig. 2 trace the influence of online access, estimated from the entire sample of articles, and illustrated for journals and subfields with the mean number of citations. Figure 2A shows the simultaneous effect of commercial and free online availability on the average age of citations. Consider a journal whose articles reference prior work that is, on average, 5.6 years old—the sample mean. If that journal's issues become available online for an additional 15 years, both commercially and for free, the average age of references would decrease to less than 4.5 years, falling by 0.088 years for each new online year available. The within-subfield models followed the same pattern, although confidence intervals were wider (tables S2 to S4).

To determine the effect of online availability on the amount of distinct research cited, I explored the relation between the distinct number of articles and journals cited in a given citation year by depth of online availability. The number of distinct articles and journals was calculated over a 20-year window, as in the previous analysis. For the average journal, 632 articles were cited each year, but this ranges widely. Because citation values are discrete and because high values concentrate within a few core journals but vary widely among the others, I modeled its relation with online availability by means of neg-

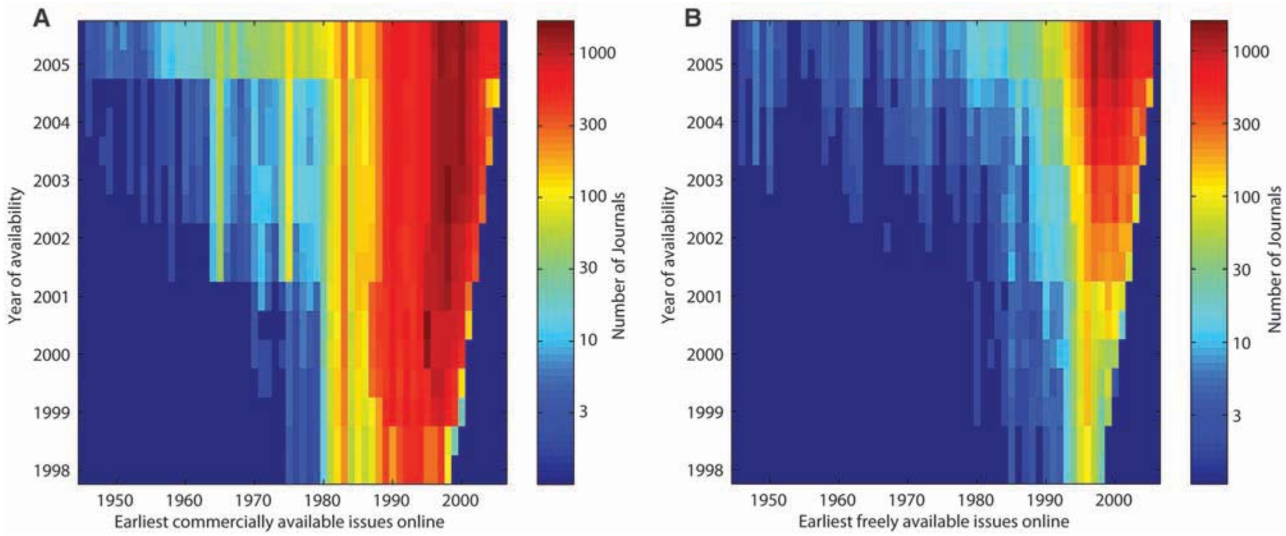


Fig. 1. Distribution of online journal availability in ISI-FSO data through (**A**) commercial subscription and (**B**) free through journal Web site. "Hot" regions of the graph correspond to journal issues just a few years behind the years in which they are available online, e.g., in 2003, more journals were commercially and freely available from 1999—about 1000 and 500, respectively—than from any other year. The figure highlights how journal issues increasingly came online from the 1940s, '50s, and '60s in 2004 and 2005.

ative binomial models (*12*). The negative binomial is a generalization of the Poisson model that allows for an additional source of variance above that due to pure sampling error. A fixed-effects specification of this model refers not to the coefficient estimates but to the "dispersion parameter," forcing the estimated variance of citations to be the same within journals or subfields, but allowing it to take on any value across them. These models were estimated with the maximum likelihood method and produced coefficient estimates that, when exponentiated, can be interpreted as the ratio of (i) the number of distinct articles cited after a 1-year increase in the electronic provision of journals over (ii) the number of articles cited without an online increase. One can subtract 1 from these ratios and multiply by 100 to obtain the percentage change of a 1-year increase in online availability on the number of distinct items cited. All models contained measures that statistically control for citation year, average number of pages, and references in citing articles.

In each subsequent year from 1965 to 2005, more distinct articles were cited from journals and subfields. The pool of published science is growing, and more of it is archived in the CI each year. Online availability, however, has not driven this trend. Figure 2B illustrates the simultaneous effect of free and online availability on the number of distinct articles cited in journals, and the number of distinct articles and journals cited in subfields. The panels portray these effects for a hypothetical journal and subfield receiving the sample mean of citations. With five additional years of free and commercial online availability, the number of distinct articles cited within journal would drop from 600 to 200; the number of articles cited within subfields would drop from 25,000 to 15,000; and the number of journals cited within subfields would drop from 19 to 16. This suggests that online availability may have reduced the number of distinct articles and journals cited below what it would have been had journals not gone online. Provision of one additional year of issues online for free associates with 14% fewer distinct articles cited.

Fewer distinct articles and journals were cited soon after they went online. Although this influenced the overall concentration of article

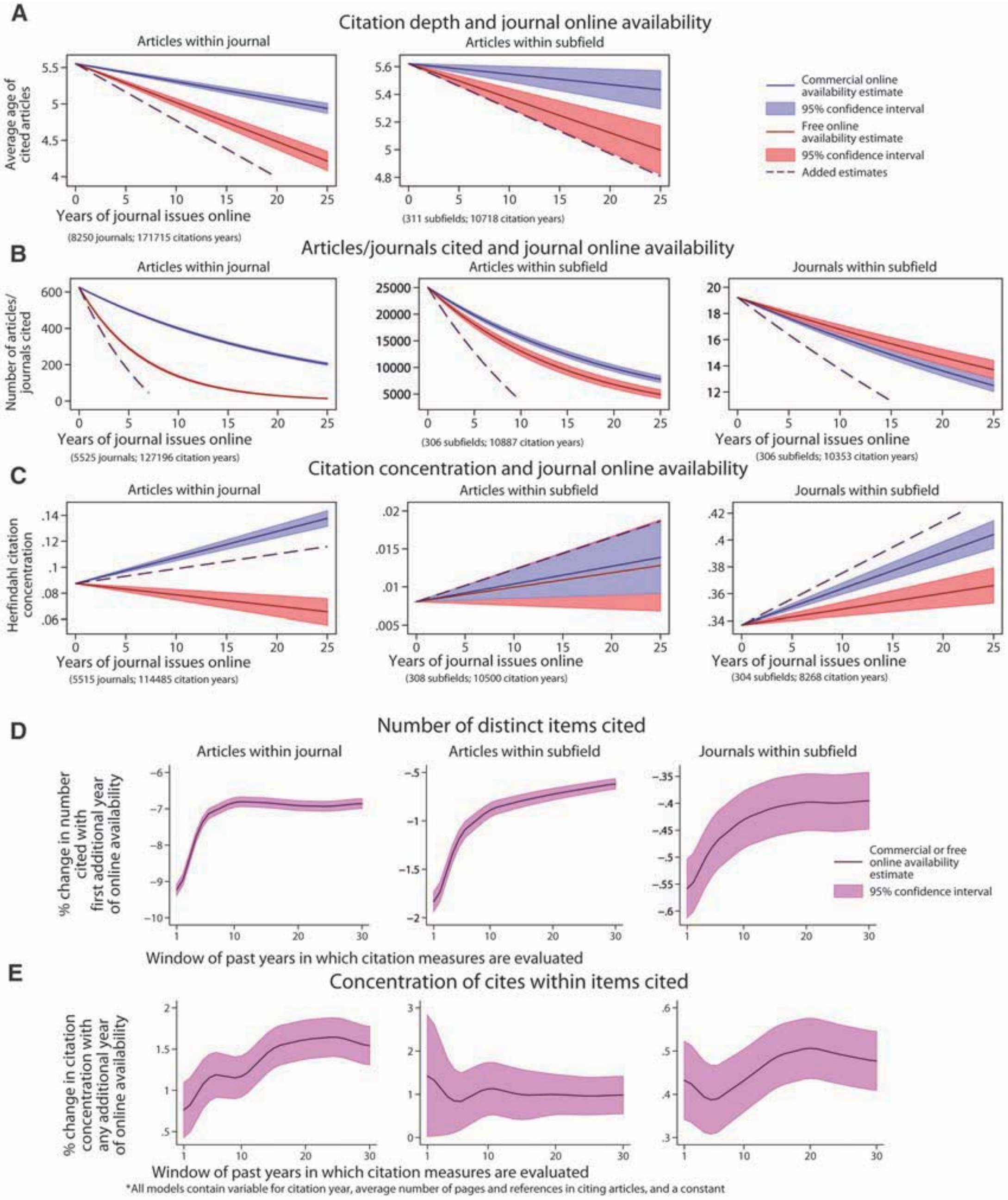


Fig. 2. Estimated influence of commercial and free online article availability (in years of journal issues available online) on (**A**) mean age of citations (based on OLS regression coefficients); (**B**) distinct number of articles and journals cited (based on exponentiated maximum likelihood negative binomial regression coefficients); and (**C**) Herfindahl concentration of citations within particular articles and journals (based on OLS regression coefficients). Each of these relations is illustrated relative to the sample mean of citation age, number, and concentration; each relation illustrated represents an underlying model that accounts for citation year, number of pages, and number of references in citing articles; the underlying citation age model also accounts for the mean weighted age of weighted title words in citing articles. Estimated percentage change, given one additional year of online availability, for (**D**) number of distinct articles and journals cited and (**E**) Herfindahl concentration within those citations, when enlarging the window in which citation measures are evaluated, from 1 to 30 years—1975 to 2005.

citations in science, it did not fully determine it. Citations may be spread more evenly over fewer articles to more broadly disperse scientific attention. To assess the degree to which online provision influences the concentration of citations to just a few articles (and journals), I computed a Herfindahl index, where $\sum_{j=1}^{n}(s_j^2)$ represents the percentage of citations s to each article j, squared and summed across journal or subfield i within the 20-year time window examined. A concentration of 1 indicates that every citation to journal i in a given year is to a single article; a concentration just less than 1 suggests a high proportion of citations pointing to just a few articles; and a concentration approaching zero implies that citations reach out evenly to a large number of articles. Herfindahl concentrations of articles cited in journals ranged from 0.0000933 to 1 in this sample, with an average of 0.088 and a wide standard deviation of 0.195. Where no articles were cited, no concentrations could be computed. Regression models were used to examine whether citation concentration to articles from the last 20 years could be attributed to depth of online availability. As in previous models, these were estimated for articles within journals and for articles and journals within subfields, by means of both commercial and free electronic provision. Citation concentrations are approximately normally distributed and the models were estimated with OLS.

Figure 2C illustrates the concurrent influence of commercial and free online provision on the concentration of citations to particular articles and journals. The left panel shows that the number of years of commercial availability appears to significantly increase concentration of citations to fewer articles within a journal. If an additional 10 years of journal issues were to go online via any commercial source, the model predicts that its citation concentration would rise from 0.088 to 0.105, an increase of nearly 20%. Free electronic availability had a slight negative effect on the concentration of articles cited within journals, but it had a marginally positive effect on the concentration of articles cited within subfields (middle panel) and appeared to substantially drive up the concentration of citations to central journals within subfields (right panel). Commercial provision had a consistent positive effect on citation concentration in both articles and journals. The collective similarity between commercial and free access for all models discussed suggests that online access—whatever its source—reshapes knowledge discovery and use in the same way. For all models, similar results were obtained when journals' presence in multiple (e.g., one, two, and three or more) commercial archives was accounted for and modeled simultaneously.

Although 20 years is not an unreasonable window of time within which to examine the effect of online availability on citations, it does not capture the trend of the effect. For example, one can imagine that online provision increases the distinct number of articles cited and decreases the citation concentration for recent articles, but hastens convergence to canonical classics in the more distant past. To explore this possibility, I performed the same analyses but calculated variables with expanding windows ranging from the last year to the last 30 years. To keep samples comparable, I estimated all models on data from 1975 (1945 plus a 30-year window) to 2005, and so the 20-year window coefficients do not correspond perfectly to the effects illustrated earlier. Estimated percentage changes in the number of articles and journals cited and the Herfindahl citation concentration within those citations were calculated as associated with a 1-year extension of online availability. These estimates and their corresponding 95% confidence intervals are graphed in Fig. 2, D and E. Increased online provision in the preceding year was associated with a decrease in the number of distinct articles cited within journals and articles and journals cited within subfields most in recent years (Fig. 2D). A 1-year change in online availability corresponded to a 9% drop in articles cited in the last year, but only a 7% drop in articles cited in the past 20 and 30 years. The pattern was the same for articles and journals within subfields (tables S2 to S4). The citation window's effect on citation concentration was not so consistent (Fig. 2E). Nevertheless, in the case of article concentrations within subfields, the Herfindahl concentration increase was highest—1.5% per year of online availability—when calculated for references to only the last year's articles.

The models presented are limited in a number of ways. For example, journals such as *Science* use Supporting Online Material for "Materials and Methods," which frequently include references not indexed by the CI. It is theoretically possible, though unlikely, that these references are to earlier or more diverse articles. Moreover, by studying only conventional journals, this study fails to capture newer scientific media like science blogs, wikis, and online outlets exploring alternative models of peer review. These new media almost undoubtedly link to extremely recent scientific developments—often through ephemeral Web links (*13*)—but they may also point to more diverse materials.

Collectively, the models presented illustrate that as journal archives came online, either through commercial vendors or freely, citation patterns shifted. As deeper backfiles became available, more recent articles were referenced; as more articles became available, fewer were cited and citations became more concentrated within fewer articles. These changes likely mean that the shift from browsing in print to searching online facilitates avoidance of older and less relevant literature. Moreover, hyperlinking through an online archive puts experts in touch with consensus about what is the most important prior work—what work is broadly discussed and referenced. With both strategies, experts online bypass many of the marginally related articles that print researchers skim. If online researchers can more easily find prevailing opinion, they are more likely to follow it, leading to more citations referencing fewer articles. Research on the extreme inequality of Internet hyperlinks (*14*), scientific citations (*15*, *16*), and other forms of "preferential attachment" (*17*, *18*) suggests that near-random differences in quality amplify when agents become aware of each other's choices. Agents view others' choices as relevant information—a signal of quality—and factor them into their own reading and citation selections. By enabling scientists to quickly reach and converge with prevailing opinion, electronic journals hasten scientific consensus. But haste may cost more than the subscription to an online archive: Findings and ideas that do not become consensus quickly will be forgotten quickly.

This research ironically intimates that one of the chief values of print library research is poor indexing. Poor indexing—indexing by titles and authors, primarily within core journals—likely had unintended consequences that assisted the integration of science and scholarship. By drawing researchers through unrelated articles, print browsing and perusal may have facilitated broader comparisons and led researchers into the past. Modern graduate education parallels this shift in publication—shorter in years, more specialized in scope, culminating less frequently in a true dissertation than an album of articles (*19*).

The move to online science appears to represent one more step on the path initiated by the much earlier shift from the contextualized monograph, like Newton's *Principia* (*20*) or Darwin's *Origin of Species* (*21*), to the modern research article. The *Principia* and *Origin*, each produced over the course of more than a decade, not only were engaged in current debates, but wove their propositions into conversation with astronomers, geometers, and naturalists from centuries past. As 21st-century scientists and scholars use online searching and hyperlinking to frame and publish their arguments more efficiently, they weave them into a more focused—and more narrow—past and present.

References and Notes

1. R. Reddy *et al.*, "Digital Libraries: Universal Access to Human Knowledge" (President's Information Technology Advisory Committee, Panel on Digital Libraries, 2001); www.nitrd.gov/pubs/pitac/pitac-dl-9feb01.pdf.
2. The report (*1*) qualifies the vision of universal access, but only by admitting that "more 'quality' digital contents" must be made available and better IT infrastructure must deliver them.
3. M. McLuhan, *Understanding Media* (McGraw-Hill, New York, 1964), chap. 1.
4. S. Black, *Libr. Resour. Tech. Serv.* **49**, 19 (2005).
5. S. L. De Groote, J. L. Dorsch, *J. Med. Libr. Assoc.* **91**, 231 (2003).
6. C. Tenopir, B. Hitchcock, S. A. Pillow, "Use and Users of Electronic Library Resources: An Overview and Analysis of

Recent Research Studies" (Council on Library and Information Resources, Washington, DC, 2003).
7. A. Friedlander, "Dimensions and Use of the Scholarly Information Environment: Introduction to a Data Set Assembled by the Digital Library Federation and Outsell, Inc." (Council on Library and Information Resources, Washington, DC, 2002); www.clir.org/pubs/reports/pub110/contents.html.
8. P. Boyce, D. W. King, C. Montgomery, C. Tenopir, *Ser. Libr.* **46**, 121 (2004).
9. C. Tenopir, D. W. King, A. Bush, *J. Med. Libr. Assoc.* **92**, 233 (2004).
10. C. Shirky, "Ontology is Overrated: Categories, Links and Tags" (Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community, Open Source, 2005); www.shirky.com/writings/ontology_overrated.html.
11. C. Manning, H. Schütz, *Foundations of Natural Language Processing* (MIT Press, Cambridge, MA, 1999).
12. J. Hausman, B. H. Hall, Z. Griliches, *Econometrica* **52**, 909 (1984).
13. R. P. Dellavalle *et al.*, *Science* **302**, 787 (2003).
14. A. L. Barabási, R. Albert, *Science* **286**, 509 (1999).
15. R. K. Merton, *Science* **159**, 56 (1968).
16. D. J. de Solla Price, *Science* **149**, 510 (1965).
17. H. A. Simon, *Biometrika* **42**, 425 (1955).
18. M. J. Salganik, P. S. Dodds, D. J. Watts, *Science* **311**, 854 (2006).
19. J. Berger, "Exploring ways to shorten the ascent to a Ph.D.," *New York Times*, 3 October 2007; www.nytimes.com/2007/10/03/education/03education.html.
20. I. Newton, *Principia* (Macmillan, New York, ed. 4, 1883) (first published in 1687).
21. C. Darwin, *The Origin of Species* (D. Appleton, New York, 1867) (first published in 1859).

22. I gratefully acknowledge research support from NSF grant 0242971, Science Citation Index data from Thompson Scientific, Inc., and *Fulltext Sources Online* data from Information Today, Inc. I also thank J. Reimer for helpful discussion and insight.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5887/395/DC1
Methods
Tables S1 to S4
References

13 September 2007; accepted 9 June 2008
10.1126/science.1150473

The Evolution and Distribution of Species Body Size

Aaron Clauset[1]* and Douglas H. Erwin[1,2]

The distribution of species body size within taxonomic groups exhibits a heavy right tail extending over many orders of magnitude, where most species are much larger than the smallest species. We provide a simple model of cladogenetic diffusion over evolutionary time that omits explicit mechanisms for interspecific competition and other microevolutionary processes, yet fully explains the shape of this distribution. We estimate the model's parameters from fossil data and find that it robustly reproduces the distribution of 4002 mammal species from the late Quaternary. The observed fit suggests that the asymmetric distribution arises from a fundamental trade-off between the short-term selective advantages (Cope's rule) and long-term selective risks of increased species body size in the presence of a taxon-specific lower limit on body size.

Most taxonomic groups show a common distribution of species body size (*1*–*3*), with a single prominent mode relatively near but not at the smallest species size (*4*) and a smooth but heavy right tail (often described as a right skew on a log-size scale) extending for several orders of magnitude (e.g., Fig. 1). This distribution is naturally related to a wide variety of other species characteristics with which body size correlates, including habitat, life history, life span (*5*), metabolism (*6*), and extinction risk (*7*). A greater understanding of the underlying constraints on, and long-term trends in, body size evolution may provide information for conservation efforts (*8*) and insight about interactions between ecological and macroevolutionary processes (*9*).

Studies of body-size distributions have suggested that the prominent mode may be indicative of a taxon-specific energetically optimal body size (*10*, *11*), which is supported by microevolutionary studies of insular species (*12*). However, evidence for Cope's rule (*1*, *13*, *14*)—the observation that species tend to be larger than their ancestors—and the fact that most species are not close to their group's predicted optimal size [among other reasons (*15*)] suggest that this theory may be flawed. Alternatively, species body sizes may diffuse over evolutionary time. If so, Cope's rule alone could cause size distributions to exhibit heavy right tails (*1*), although size-dependent speciation or extinction rates (*2*, *9*, *16*) or size-neutral diffusion near a taxon-specific lower limit on body size (*17*) could also produce a similar shape. Furthermore, different mechanisms may drive body-size evolution on spatial and temporal scales (*3*), and the importance of interspecific competition to the macroevolutionary dynamics of species body size is not known.

We developed a generalized diffusion model of species body-size evolution, in which the size distribution is the product of three macroevolutionary processes (Fig. 1). We combine these processes, each of which has been independently studied (*1*, *2*, *17*, *18*), in a single quantitative framework, estimate its parameters from fossil data on extinct terrestrial mammals from before the late Quaternary (*19*, *20*), and determine whether this model, or simpler variants, can reproduce the sizes of the 4002 known extant and extinct terrestrial mammal species from the late Quaternary (Recent species) (*21*, *22*).

This model makes three assumptions: (i) Species size varies over evolutionary time as a cladogenetic multiplicative diffusion process (*1*, *17*); the size of a descendant species x_D is the product of a stochastic growth factor λ and its ancestor's size x_A, that is, $x_D = \lambda x_A$. For each speciation event, a new λ is drawn from the distribution $F(\lambda)$, which models the total influence on species size changes from all directions. A bias toward larger sizes (Cope's rule) appears as a positive average log-change to size $\langle \log \lambda \rangle > 0$, and may depend on the ancestor's size. (ii) Species body size is restricted by a taxon-specific lower limit $x_{\min}$ (*6*, *23*), which we model by requiring that $F(\lambda < x_{\min}/x_A) = 0$, that is, the largest possible decrease in size for a particular speciation event is $\lambda = x_{\min}/x_A$. In our computer simulations, time proceeds in discrete steps. At each step, exactly one new species is produced, which is the descendant of a randomly selected species. (iii) Every species independently becomes extinct with probability $p_e(x)$, which increases monotonically with size. A schematic of the model is shown in Fig. 2A. [For technical details, see (*24*).]

To make this model appropriately realistic, we estimated the form of each process from fossil data. The lower limit on mammalian body size is near 2 g, close to the size of both the Etruscan shrew (*Suncus etruscus*) and the bumblebee bat (*Craseonycteris thonglongyai*). Fossil evidence suggests that this limit has existed since at least the Cretaceous-Tertiary boundary (*19*, *20*, *25*). Further, a limit in this vicinity is supported by both

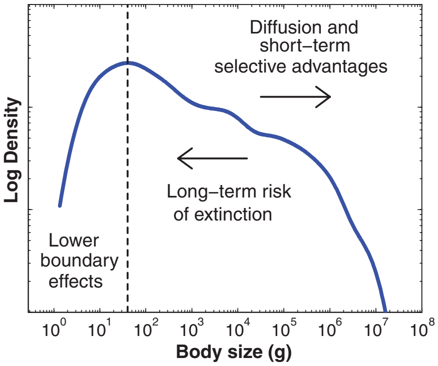


Fig. 1. Smoothed species body-size distribution of 4002 Recent terrestrial mammals [data from (*21*)], showing the three macroevolutionary processes that shape the relative abundances of different sizes. The left tail of the distribution is created by diffusion in the vicinity of a taxon-specific lower limit near 2 g, whereas the long right tail is produced by the interaction of diffusion over evolutionary time (including trends like Cope's rule) and the long-term risk of extinction from increased body size.

---

[1] Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA. [2] Department of Paleobiology, MRC-121, National Museum of Natural History, Post Office Box 37012, Washington, DC 20013–012, USA.

*To whom correspondence should be addressed. E-mail: aaronc@santafe.edu

experimental (*23*) and theoretical work (*6*) on mammalian metabolism.

Away from this limit, mammalian body-size evolution is governed mainly by diffusion with a bias (Cope's rule) (*14*, *26*), whereas its evolution near the lower limit is likely constrained by the need for relatively specialized morphological structures (*1*). We expect this latter effect to appear in fossil data as a systematic intensification of Cope's rule for very small-bodied species, that is, increased $\langle \log \lambda \rangle$ as $x_A \rightarrow x_{\min}$. From ancestor-descendant size data for 1106 extinct North American terrestrial mammals (*20*), we estimated and compared three models of the distribution $F(\lambda)$ as a function of ancestor size, including the model suggested by Alroy (*14*), which predicts a moderately bimodal distribution in body sizes. Of these, a piecewise model (Fig. 2B), with no effective optimal body size, has the best empirical support [model selection by likelihood ratio test and Bayesian information criterion (*24*)]. This model includes both a strengthening of Cope's rule for small-bodied species ($x \lesssim 32$ g) and a small but uniformly positive bias for larger species, resulting in an average body-size growth of 4.1 ± 1.0% between ancestors and their descendants ($\langle \log \lambda \rangle = 0.04 \pm 0.01$).

This result supports the existence of short-term selective advantages for increased species body size—for example, better tolerance of resource fluctuations, better thermoregulation, and better predator avoidance (*5*)—but also implies a more nuanced view: Small-bodied species exhibit even greater selective advantages from increased size, for example, because of greater morphological flexibility.

Empirical estimates of extinction rates (or, equivalently, speciation rates) as functions of body size are uncertain (*27*) because of the bias and incompleteness of the fossil record. We partly control for this uncertainty by using a simplistic model of extinction risk $p_e(x)$, largely estimated from the data, where extinction occurs independently, with a probability calculated only from the species' size. We specified a basal extinction rate β by assuming that the number of Recent terrestrial mammal species is close to a putative carrying capacity. We then let extinction risk per unit time increase logarithmically with body size (*24*, *28*). This model leaves only the rate ρ by which risk increases with size as a free parameter, which was chosen by minimizing the statistical distance between the simulated and empirical distributions (*24*).

Inserting these three processes, as estimated above, into our computer model, we found that the model accurately predicted the distribution of Recent terrestrial mammal sizes over its seven orders of magnitude (Fig. 3A) and was particularly accurate for small-bodied species ($x < 80$ g). Our sensitivity analysis further indicated that this prediction was highly robust to variations in most of the estimated parameters, but highly sensitive to the location of the lower limit on body size. The estimated value of $x_{\min} \approx 2$ g, however, is the most strongly supported of all model parameters. Thus, even large revisions to the other parameter estimates are unlikely to change our general conclusions (*24*). Also, although a range of ρ values produced size distributions that were statistically close to the empirical distribution, the model predicts a particular extinction risk curve (fig. S4B) that could be tested with appropriate empirical data.

To further discriminate among alternative explanations for the species size distribution, we tested simpler diffusion models, each with parameters estimated from fossil data (*24*), including (i) unbiased diffusion with a lower boundary, (ii) Cope's rule with size-dependent extinction, (iii) Cope's rule alone, (iv) size-dependent extinction alone, and (v) a version of the full model that omits the increased bias for small-bodied species ($x \lesssim 32$ g). We found that these models all predicted size distributions that differed, sometimes dramatically so, from the empirical distribution (Fig. 3, B and C, and figs. S9 and S10). Additionally, we found that a positive bias $\langle \log \lambda \rangle > 0$ for large-bodied species is not necessary if the extinction risk increases less quickly (*24*). These results support the inclusion of a fundamental lower limit, the diffusion of species size, and an increasing risk of extinction with size, as well as an increased bias toward larger sizes for small-bodied species ($x \lesssim 32$ g).

Thus, the shape of a body-size distribution can be interpreted in the context of these three macroevolutionary processes. An intermediate lo-

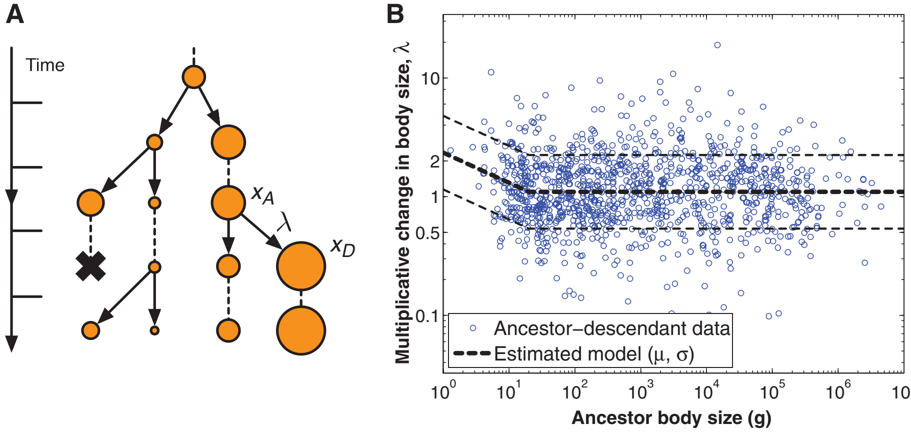


Fig. 2. (**A**) A schematic illustrating a simple cladogenetic diffusion model of species body-size evolution, where the size of a descendant species x_D is related to its ancestor's size x_A by a multiplicative factor λ. (**B**) Empirical data on 1106 changes in North American mammalian body size [data from (*20*)], as a function of ancestor size, overlaid with the estimated model of within-lineage changes, where the average log-change $\langle \log \lambda \rangle$ varies piecewise as a function of body size (*24*).

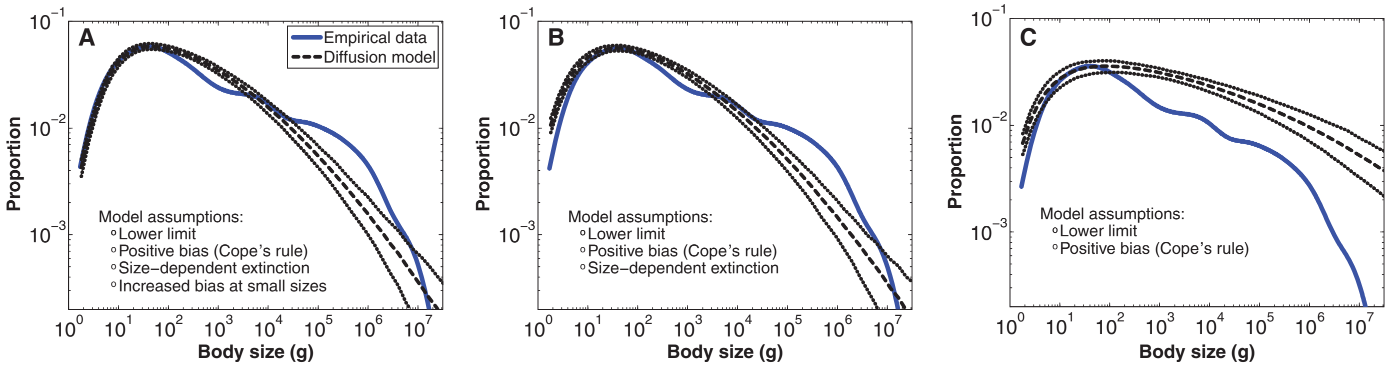


Fig. 3. Simulated distributions of species body size (central tendency ± 95% confidence intervals from 1000 repetitions; all model parameters estimated as described in the text) and the empirical distribution of Recent terrestrial mammals. (**A**) The model described in the text. (**B**) The same model as in (A) but with a bias $\langle \log \lambda \rangle$ that is independent of size. (**C**) The same model as in (B) but with an extinction risk that is independent of size. [For details and additional results, see (*24*).]

cation for the distribution's mode (40 g for terrestrial mammals) is mainly caused by diffusion in the vicinity of the physiological lower limit on body size, which prevents the smallest species from being the most abundant. A heavy right tail is then caused primarily by diffusion in the presence of extinction risks that increase weakly with size ($\rho > 0$). For mammals, the within-lineage tendency toward increased size (Cope's rule, $\langle \log \lambda \rangle > 0$) shifts the mode toward slightly larger sizes and slightly increases the heaviness of the right tail.

Under different conditions, these processes produce markedly different body size distributions. For instance, a long left tail extending toward small-bodied species would indicate that the risk of extinction decreases with larger size ($\rho < 0$). Similarly, a more symmetric distribution would indicate both that extinction rates are relatively size-independent ($\rho \approx 0$) and that changes to body size convey few selective advantages ($\langle \log \lambda \rangle \approx 0$). Although a suitable body-size distribution is not currently available for dinosaurs [but see (*29*)], evidence suggests that it may be more symmetric than for mammals. The right-skewed distribution's ubiquity, such as for insects and birds (*1*, *2*), suggests that such circumstances are rare and that the mammalian distribution represents the norm.

This model omits explicit mechanisms for many canonical ecological and microevolutionary processes, including the impact of interspecific competition, geography, predation, population dynamics, and size variation between speciation events (anagenetic evolution), which suggests that their contributions to the systematic or large-scale character of species body-size distributions can be compactly summarized by the values of certain model parameters, for example, the strength of Cope's rule $\langle \log \lambda \rangle$ or the manner in which extinction risk increases with body size ρ. Some aspects of the body-size distribution, however, are not explained by this model, such as the slight overabundance of terrestrial mammal species around 300 kg and the slight underabundance around 1 kg (Fig. 3A). Whether such deviations can be attributed to phylogenetically correlated speciation and extinction events is an open question. A more thorough examination of these macroevolutionary processes may explain their particular form and origin, as well as why body size is weakly correlated with increased extinction rates (or decreased speciation rates), why physiological lower limits on body size exist and are conserved within taxonomic groups, and why some groups exhibit macroevolutionary trends but others do not.

References and Notes

1. S. M. Stanley, *Evol. Int. J. Org. Evol.* **27**, 1 (1973).
2. J. Kozlowski, A. T. Gawelczyk, *Funct. Ecol.* **16**, 419 (2002).
3. C. R. Allen *et al.*, *Ecol. Lett.* **9**, 630 (2006).
4. K. P. Dial, J. M. Marzluff, *Ecology* **69**, 1620 (1988).
5. J. H. Brown, *Macroecology* (Univ. of Chicago Press, Chicago, 1995).
6. G. B. West, W. H. Woodruff, J. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2473 (2002).
7. M. Cardillo *et al.*, *Science* **309**, 1239 (2005).
8. D. O. Fisher, I. P. F. Owens, *Trends Ecol. Evol.* **19**, 391 (2004).
9. S. M. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 646 (1975).
10. K. P. Sebens, *Annu. Rev. Ecol. Syst.* **18**, 371 (1987).
11. J. H. Brown, P. A. Marquet, M. L. Taper, *Am. Nat.* **147**, 1092 (1996).
12. M. Lomolino, *Am. Nat.* **125**, 310 (1985).
13. C. Deperet, *The Transformations of the Animal World* (D. Appleton, New York, 1909).
14. J. Alroy, *Science* **280**, 731 (1998).
15. J. Kozlowski, *Funct. Ecol.* **16**, 540 (2002).
16. L. Van Valen, *Evol. Int. J. Org. Evol.* **29**, 87 (1973).
17. D. W. McShea, *Evol. Int. J. Org. Evol.* **48**, 1747 (1994).
18. M. L. McKinney, in *Evolutionary Trends*, K. J. McNamara, ed. (Univ. of Arizona Press, Tucson, AZ, 1990), pp. 75–118.
19. M. Fortelius, Neogene of the Old World Database of Fossil Mammals (NOW) (2003). University of Helsinki, NOW public release 030717, www.helsinki.fi/science/now.
20. J. Alroy, North American Fossil Mammal Systematics Database (2008). Paleobiology Database Online Systematics Archive 3, http://paleodb.org.
21. F. A. Smith *et al.*, *Ecology* **84**, 3403 (2003).
22. F. A. Smith *et al.*, Macroecological Database of Mammalian Body Mass, MOM v 3.6.1 (2007).
23. O. P. Pearson, *Science* **108**, 44 (1948).
24. Materials and methods are available as supporting material on *Science* Online.
25. F. A. Smith *et al.*, *Am. Nat.* **163**, 672 (2004).
26. B. Van Valkenburgh, X. Wang, J. Damuth, *Science* **306**, 101 (2004).
27. D. Ludwig, *Ecol. Appl.* **6**, 1067 (1996).
28. L. H. Liow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6097 (2008).
29. M. T. Carrano, *Amniote Paleobiology*, M. T. Carrano, T. J. Gaudin, R. W. Blob, J. R. Wible, Eds. (Univ. of Chicago Press, Chicago, 2006), pp. 225–268.

30. A.C. is grateful to A. Boyer, J. Dunne, J. Ladau, B. Olding, C. Shalizi, and J. Wilkins for helpful conversations. We thank J. Alroy, A. Boyer, and F. Smith for sharing data. Supported in part by the Santa Fe Institute and the Computer Science Department at the University of New Mexico.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/399/DC1
SOM Text
Figs. S1 to S10
Tables S1 and S2
References

10 March 2008; accepted 22 May 2008
10.1126/science.1157534

Four-jointed Is a Golgi Kinase That Phosphorylates a Subset of Cadherin Domains

Hiroyuki O. Ishikawa,[1] Hideyuki Takeuchi,[2] Robert S. Haltiwanger,[2] Kenneth D. Irvine[1]*

The atypical cadherin Fat acts as a receptor for a signaling pathway that regulates growth, gene expression, and planar cell polarity. Genetic studies in *Drosophila* identified the *four-jointed* gene as a regulator of Fat signaling. We show that *four-jointed* encodes a protein kinase that phosphorylates serine or threonine residues within extracellular cadherin domains of Fat and its transmembrane ligand, Dachsous. Four-jointed functions in the Golgi and is the first molecularly defined kinase that phosphorylates protein domains destined to be extracellular. An acidic sequence motif (Asp-Asn-Glu) within Four-jointed was essential for its kinase activity in vitro and for its biological activity in vivo. Our results indicate that Four-jointed regulates Fat signaling by phosphorylating cadherin domains of Fat and Dachsous as they transit through the Golgi.

The Fat and Hippo signaling pathways intersect at multiple points and influence growth and gene expression through regulation of the transcriptional coactivator Yorkie (*1*–*8*). Fat signaling also influences planar cell polarity (PCP) (*9*). Fat acts as a transmembrane receptor, and is a large (5147 amino acids) atypical cadherin protein, with 34 extracellular cadherin domains (Fig. 1A) (*10*). Dachsous (Ds) is also a large (3503 amino acids) transmembrane protein with multiple cadherin domains (Fig. 1A) (*11*) and is a candidate Fat ligand because it appears to bind Fat in a cultured cell assay (*12*), acts non–cell autonomously to influence Fat pathway gene expression (*2*, *13*,), and acts genetically upstream of *fat* in the regulation of PCP (*14*). A second protein, Four-jointed (Fj), also acts non–cell autonomously to influence Fat pathway gene expression and acts genetically upstream of *fat* in the regulation of PCP (*2*, *13*–*15*). However, Fj is a type II transmembrane protein that functions in the Golgi (*16*, *17*). Thus, Fj might influence Fat signaling by posttranslationally modifying a component of the Fat pathway.

To investigate the possibility of modification of Fat or Ds, we coexpressed FLAG epitope–tagged fragments of their extracellular domains together with Fj in cultured *Drosophila* S2 cells.

---

[1] Howard Hughes Medical Institute, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ 08854, USA. [2] Department of Biochemistry and Cell Biology, Institute for Cell and Developmental Biology, Stony Brook University, Stony Brook, NY 11794, USA.

*To whom correspondence should be addressed. E-mail: irvine@waksman.rutgers.edu

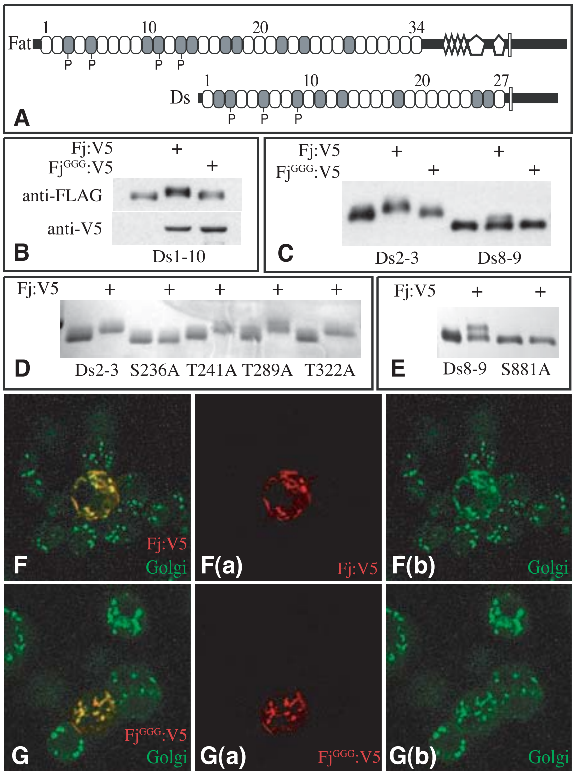


Fig. 1. Four-jointed–dependent mobility shift of cadherin domain polypeptides. (**A**) Schematics of Fat and Ds Cadherin domains (ovals), epidermal growth factor–like domains (diamonds), LamG domains (pentagons), and transmembrane domains (rectangles) are indicated. Shading identifies cadherin domains with a Ser or Thr as the seventh amino acid; "P" identifies cadherin domains at which Fj-mediated phosphorylation was detected. (**B**, **C**, and **E**) Western blots (anti-FLAG) of Ds cadherin domains isolated from the medium of S2 cells; cells were cotransfected with plasmids to express Fj:V5 or Fj^{GGG}:V5, as indicated. (B) Ds1-10:FLAG. Fj expression is indicated by anti-V5 staining. (C) Ds2-3:FLAG and Ds8-9:FLAG. (**D**) Coomasie stain of Ds2-3:FLAG and point mutants immunoprecipitated from the medium of transfected S2 cells, where indicated cells express Fj:V5. (E) Ds8-9:FLAG and Ds8-9^{S881A}:FLAG mutant. (**F** and **G**) S2 cells transfected to express Fj:V5 (F) or Fj^{GGG}:V5 (G) and stained with anti-V5 (red) and anti-Golgi (green); panels to the right show individual stains.

When the first 10 cadherin domains of Ds (Ds1-10) were coexpressed with Fj, a shift in mobility was observed (Fig. 1B). A common posttranslational modification of secreted and transmembrane proteins as they pass through the Golgi is glycosylation. Most glycosyltransferases contain a conserved sequence motif, Asp-X-Asp (DXD; X, any amino acid) (*18*), which is essential for their activity. Because a related sequence motif [Asp-Asn-Glu (DNE) at amino acids 490 to 492] is present in Fj and its vertebrate homologs (*19*, *20*), we created a mutant form of Fj in which DNE was changed to GGG (Fj^{GGG}; G, glycine). The expression levels and Golgi localization of Fj^{GGG} appear normal (Fig. 1, B, F, and G), but Fj^{GGG} expression did not shift Ds1-10 mobility (Fig. 1B).

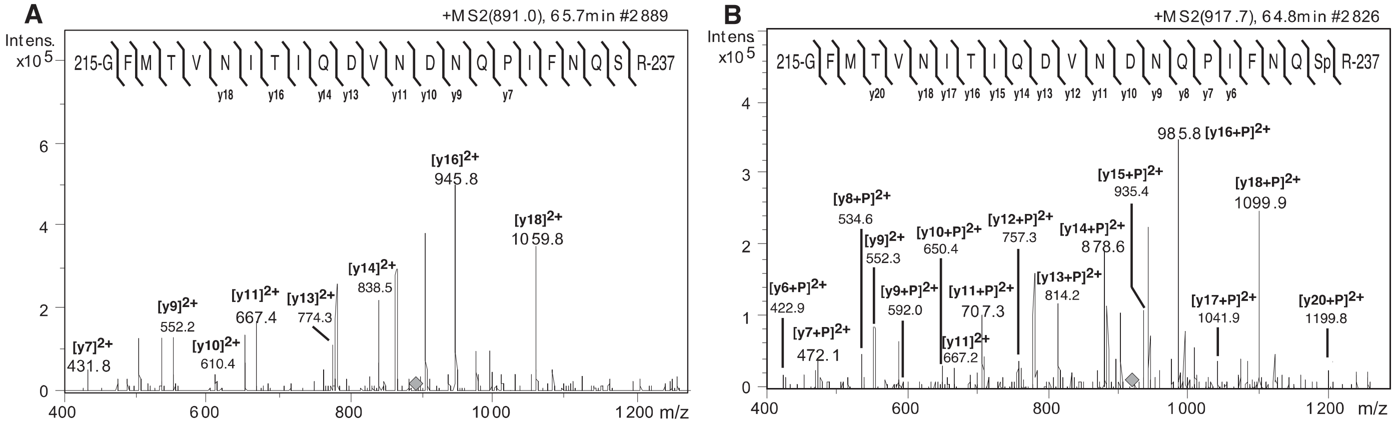
To identify modified cadherin domains, we expressed smaller fragments of Ds1-10. The smallest fragments whose mobility was shifted in cells expressing Fj were two-cadherin–domain polypeptides: Ds2-3, Ds5-6, and Ds8-9 (Fig. 1C and fig. S1C). Ds2-3 and Ds5-6 appeared to be stoichiometrically modified in cells expressing Fj, whereas Ds8-9 was only partially modified. Fat4-5 was also partially shifted by Fj coexpression (fig. S1C). The mobility shifts of these two-cadherin–domain polypeptides were not observed with Fj^{GGG} (Fig. 1C and fig. S1B). To identify potential sites of modification, we aligned their sequences (fig. S1E). This identified four sites at which a Ser or Thr residue was conserved, whose hydroxyl groups could potentially be sites of posttranslational modification. To evaluate their influence, we mutated each in turn to Ala within the Ds2-3 polypeptide. Three of the four mutants had no effect; however, one, Ds2-3^{S236A} (mutation of Ser^{236} to Ala), completely eliminated the Fj-dependent mobility shift (Fig. 1D). Introduction of an analogous mutation into Ds8-9 also eliminated its mobility shift (Fig. 1E). Thus, a Ser reside at a specific location within the second of the two cadherin domains was essential for the Fj-dependent mobility shift. This amino acid was a Ser in each of these dicadherin domains, but Thr was also compatible with the Fj-dependent modification (fig. S1D). In a structurally solved cadherin domain (*21*), this Ser is the seventh amino acid and predicted to be located on the surface near the middle of the cadherin domain (fig. S2).

To identify posttranslational modifications associated with this mobility shift, we purified Ds2-3 from S2 cells expressing or not expressing Fj, digested the proteins with trypsin, and analyzed the resulting peptides by mass spectrometry. One peptide from Fj-expressing cells was stoichiometrically shifted by 80 daltons relative to the same peptide from cells not expressing Fj, and it also eluted earlier on high-performance liquid chromatography (HPLC) (Fig. 2, A and B, and fig. S3A). Mass and tandem mass spectrometry (MS/MS) fragmentation patterns identified this peptide as amino acids 215 to 237 of Ds and refined the site of modification to within amino acids 232 to 237 (Fig. 2, A and B). The mass of the equivalent peptide from Ds2-3^{S236A} was not altered by Fj expression (fig. S3, D and E). We identified most of the peptides corresponding to Ds2-3 cadherin domains, and none of the others were detectably modified in cells expressing Fj (fig. S3, A to C). Thus, the Fj-dependent modification of Ds2-3 comprises an addition of 80 daltons, which is attached to Ser^{236}. An 80-dalton mass does not correspond to that of any known glycans, but does correspond to the mass associated with addition of a phosphate group. Incubation of Fj-modified Ds fragments with either calf intenstinal alkaline phosphatase (CIP) or Antarctic phosphatase (AnP) reversed the Fj-dependent mobility shifts of Ds2-3, Ds8-9, and Fat 4-5 (Fig. 2, C and D). Thus, Ds and Fat cadherin domains are subject to Fj-dependent phosphorylation at a specific Ser residue.

To investigate whether Fj itself has kinase activity, we purified a secreted, epitope-tagged Fj (sFj:V5) from the medium of cultured S2 cells (fig. S4). Purified sFj:V5 was then incubated with affinity-purified Ds2-3 and [γ-^{32}P]ATP (adenosine 5′-triphosphate) in buffer. Transfer of ^{32}P onto Ds2-3 was observed in the presence of sFj, but not in its absence, and not when sFj^{GGG} was used as the enzyme (Fig. 3A). Moreover, Ds2-3^{S236A} was not detectably phosphorylated by sFj. We also characterized the activity of Fj expressed in a heterologous system by expressing a glutathione S-transferase:Fj (GST:Fj) fusion protein in *Escherichi coli* and partially purifying it on glutathione beads. GST:Fj, but not GST:Fj^{GGG}, catalyzed the transfer of ^{32}P onto Ds2-3 (Fig. 3B). Thus, Fj is a protein kinase.

The generic kinase substrates myelin basic protein and casein were not detectably phosphorylated by sFj (fig. S4C). Thus, Fj appears to have a limited substrate specificity. Only a few proteins have been identified as being phosphorylated in the secretory pathway, and none of the responsible kinase(s) have been molecularly identified. A Golgi kinase activity, referred to as Golgi casein kinase, preferentially phosphorylates Ser or Thr residues within a S/T-X-E/D/S(Phos) consensus sequence (*22*, *23*). Because Fj does not phosphorylate casein (fig. S4C), and the Ser residues within cadherin domains targeted by Fj do not conform to Golgi casein kinase sites (fig. S6B), Fj is not Golgi casein kinase. Fj autophosphorylation was detected, but this reaction was weak compared to phosphorylation of Ds2-3 (Fig. 3, A and B). The autophosphorylation reaction is apparently unimolecular, because GST:Fj and sFj:V5 did not phosphorylate each other and the fraction of Fj phosphorylated was independent of concentration (fig. S5).

Some cadherin domain polypeptides that include a Ser as the seventh amino acid were not detectably shifted, but the mobility shift on Ds2-3 might reflect a conformational effect. To examine the ability of Fj to phosphorylate other cadherin domains, we performed in vitro kinase reactions with [γ-^{32}P]ATP. This identified phosphorylation sites on polypeptides that were not gel shifted, including Fat2-3, Fat10-11, and Fat12-13 (fig S6). The in vitro kinase reactions



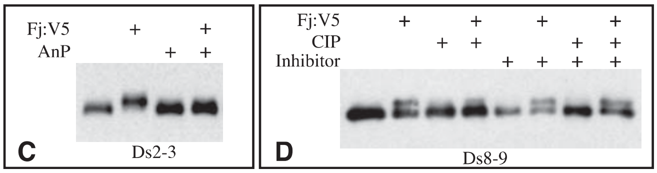


Fig. 2. Four-jointed–dependent phosphorylation of Ds and Fat cadherin domains. (**A**) MS/MS fragmentation of a tryptic peptide from Ds2-3:FLAG isolated from S2 cells, eluting on HPLC at 65.7 min (fig. S3A). The parental ion is triply charged [mass/charge ratio (*m/z*) = 891.0, gray diamond]. The mass and fragmentation pattern identify this as amino acids 215 to 237; the positions of identified *y* ions from peptide fragmentation are indicated. (**B**) MS/MS fragmentation of a tryptic peptide from Ds2-3:FLAG isolated from Fj:V5-expressing S2 cells, eluting on HPLC at 64.8 min (parent ion *m/z* = 917.7, gray diamond). The positions of identified *y* ions from peptide fragmentation are indicated. (**C**) Western blot on Ds2-3:FLAG isolated from S2 cell media; indicated cells expressed Fj:V5 and/or were treated with Antarctic phosphatase (AnP). (**D**) Western blot on Ds8-9:FLAG isolated from S2 cell media; the indicated cells expressed Fj:V5, and/or were treated with calf intestinal alkaline phosphatase (CIP), and/or were treated with phosphatase inhibitor cocktail.

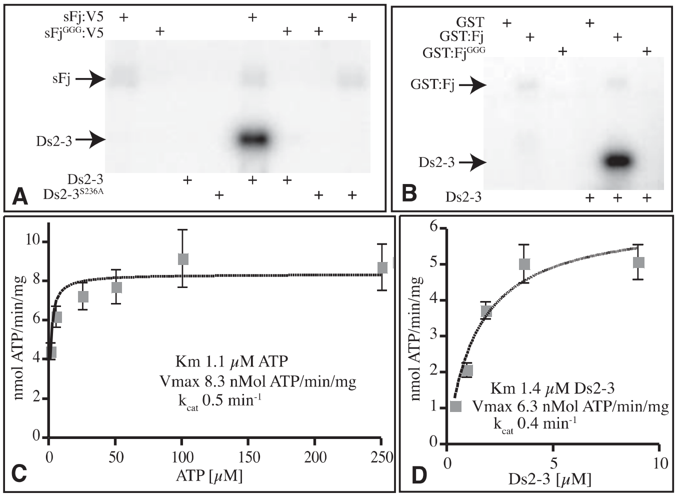


Fig. 3. Kinase activity of Four-jointed. (**A** and **B**) Autoradiograms of protein gels displaying the products of in vitro kinase reactions with the indicated enzymes (top) and substrates (bottom). The positions of Ds2-3:FLAG and sFj:V5 or GST:Fj are indicated. (A) Phosphorylation of Ds2-3, but not Ds2-3^{S236A}, by Fj in vitro. (B) Kinase activity of GST:Fj isolated from *E. coli*. (**C**) Reaction velocity versus ATP concentration, with 2.5 μg of Ds2-3:FLAG; 10 ng of sFj:V5′ and 1, 5, 25, 50, 100, or 250 μM ATP for 15 min. (**D**) Reaction velocity versus Ds2-3 concentration, with 500 μM ATP; 10 ng of sFj:V5; and 0.36, 0.90, 1.8, 3.6, or 9.0 μM Ds2-3:FLAG for 15 min. Values in (C) and (D) are the mean ± SEM from four experiments.

also identified differences in the efficiency with which different cadherin domains were phosphorylated by Fj, with Ft3, Ds3, and Ds6 being the best substrates (fig. S6).

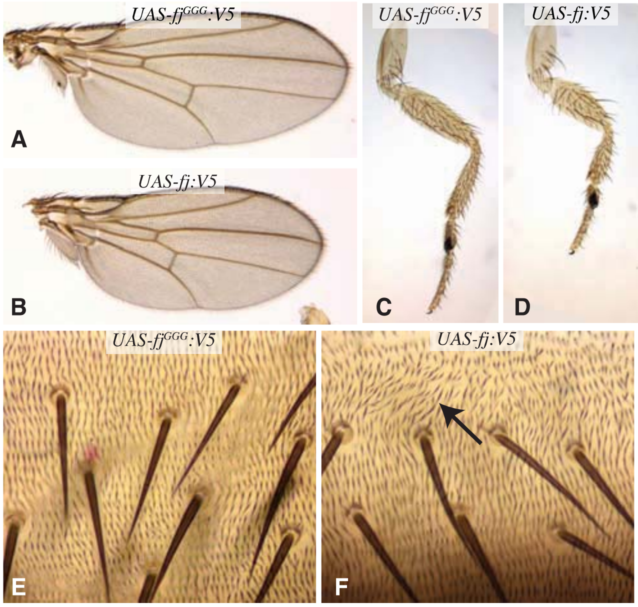
If we take the presence of a Ser or Thr at the seventh amino acid of a cadherin domain as the minimal requirement for Fj-mediated phosphorylation, there are nine potential sites in Ds and 11 in Fat (Fig. 1A). However, Fat10, Ds2, Ds11, Ds13, and Ds18 were not detectably phosphorylated, despite the presence of Ser or Thr at this position (fig. S6). Presumably, there are other structural features important for recognition by Fj. This was also emphasized by the detection of phosphorylation of our Ds2-3 polypeptide, but not our Ds3-4 polypeptide, even though both contain Ser^{236}. Our dicadherin constructs were based on published annotations (*11*), but in comparing Ds cadherin domains to structurally solved cadherin domains (*21*), we realized that these misposition the intercadherin domain boundary, and consequently these constructs lacked three amino acids of the first cadherin domain. Addition of these amino acids, together with the intercadherin domain linker sequence, enabled phosphorylation of a Ds3 single–cadherin domain construct (fig. S6D).

Parameters of Fj kinase activity were examined with the use of purified sFj:V5 and Ds2-3:FLAG. The reaction rate was linear over short times and dependent on enzyme and substrate concentrations (Fig. 3 and fig. S7). The turnover number (k_{cat}) was 0.5 min^{-1}, confirming that Fj acts catalytically. Fj exhibited a strong preference for Mn^{2+} as a cofactor and was active over a wide pH range (fig. S7). The estimated Michaelis constant (K_m) for Ds2-3 was 1.4 μM (Fig. 3D), and the K_m for ATP was ~1 μM (Fig. 3C). [γ-^{32}P]Guanosine triphosphate (GTP) could also be used as a phosphate donor, but was several-fold less effective than ATP (fig. S7E). In a competition assay, unlabeled ATP efficiently competed with [γ-P^{32}]ATP, but GTP, cytidine triphosphate, thymidine triphosphate, and uridine triphosphate did not (fig. S7F). Thus, ATP is the preferred phosphate donor.

A weak similarity between Fj and the bacterial kinase HipA, and between Fj and the mammalian lipid kinase phosphatidylinositol 4-kinase II (PI4KII), has been suggested previously on the basis of bioinformatic analyses in which HipA or PI4KII were used as the starting point for PSI-BLAST searches (*24*, *25*). Asp residues play critical roles in catalysis and in the coordination of Mg^{2+} in these and other kinases (*25*, *26*), and the loss of Fj kinase activity associated with mutation of the conserved DNE motif is thus consistent with the inference that Fj is related to other kinases. A single Fj ortholog, *Fjx1*, is present in a range of vertebrate species, including humans (*19*).

To investigate the biological requirement for Fj kinase activity, we assayed the catalytically inactive *fj^{GGG}* mutant in vivo. A V5 epitope–tagged form of this gene was expressed in transgenic *Drosophila*. At the same time, we constructed V5-tagged wild-type *fj*. To ensure that

Fig. 4. Requirement for Four-jointed kinase activity in vivo. (**A** to **F**) Portions of adult flies, expressing either wild-type *fj:V5* or the fj^{GGG}:V5 mutant. (A) Wing from *tub-Gal4 UAS-*fj^{GGG}*:V5*. (B) Wing from *tub-Gal4 UAS-fj:V5*. (C) Leg from *tub-Gal4 UAS-*fj^{GGG}*:V5*. (D) Leg from *tub-Gal4 UAS-fj:V5*. (E) Abdomen from *tub-Gal4 UAS-*fj^{GGG}*:V5* animal. (F) Abdomen from *tub-Gal4 UAS-fj:V5* animal with disturbed hair polarity (arrow).



both forms were expressed in similar amounts, we used site-specific integration to insert transgenes at the same chromosomal location (*27*). Immunostaining confirmed that Fj^{GGG}:V5 and Fj:V5 both exhibited normal Golgi localization and were expressed in similar amounts (fig. S8). Uniform overexpression of *fj* reduces the growth of legs and wings and interferes with normal PCP (*13–15*, *28*). Fj:V5 exhibited phenotypes consistent with previous studies, but Fj^{GGG}:V5 was completely inactive (Fig. 4). Thus, mutation of the DNE motif in Fj abolishes its biological activity.

The identification of Fj's cadherin domain kinase activity provides a biochemical explanation for the influence of Fj on Fat signaling and supports a model in which Fj directly phosphorylates Fat and Ds as they transit through the Golgi to influence their activity, presumably by modulating interactions between their cadherin domains. Because there was a substantial difference in the efficiency with which individual cadherin domains could be modified by Fj, both in cell-based and in vitro assays, it is also possible that differences in the extent of Fat and Ds phosphorylation normally occur in vivo and might differentially modify their binding or activity.

References and Notes

1. F. C. Bennett, K. F. Harvey, *Curr. Biol.* **16**, 2101 (2006).
2. E. Cho *et al.*, *Nat. Genet.* **38**, 1142 (2006).
3. E. Silva, Y. Tsatskis, L. Gardano, N. Tapon, H. McNeill, *Curr. Biol.* **16**, 2081 (2006).
4. M. Willecke *et al.*, *Curr. Biol.* **16**, 2090 (2006).
5. Y. Feng, K. D. Irvine, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20362 (2007).
6. J. Dong *et al.*, *Cell* **130**, 1120 (2007).
7. J. Huang, S. Wu, J. Barrera, K. Matthews, D. Pan, *Cell* **122**, 421 (2005).
8. H. Oh, K. D. Irvine, *Development* **135**, 1081 (2008).
9. H. Strutt, D. Strutt, *Bioessays* **27**, 1218 (2005).
10. P. A. Mahoney *et al.*, *Cell* **67**, 853 (1991).
11. H. F. Clark *et al.*, *Genes Dev.* **9**, 1530 (1995).
12. H. Matakatsu, S. S. Blair, *Development* **131**, 3785 (2004).
13. E. Cho, K. D. Irvine, *Development* **131**, 4489 (2004).
14. C. H. Yang, J. D. Axelrod, M. A. Simon, *Cell* **108**, 675 (2002).
15. J. Casal, G. Struhl, P. Lawrence, *Curr. Biol.* **12**, 1189 (2002).
16. H. Strutt, J. Mundy, K. Hofstra, D. Strutt, *Development* **131**, 881 (2004).
17. J. L. Villano, F. N. Katz, *Development* **121**, 2767 (1995).
18. C. A. Wiggins, S. Munro, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7945 (1998).
19. R. Rock, A. C. Heinrich, N. Schumacher, M. Gessler, *Dev. Dyn.* **234**, 602 (2005).
20. K. Yamaguchi, J. Parish, K. Akita, P. Francis-West, *Dev. Dyn.* **235**, 3085 (2006).
21. S. D. Patel *et al.*, *Cell* **124**, 1255 (2006).
22. B. Christensen, M. S. Nielsen, K. F. Haselmann, T. E. Petersen, E. S. Sorensen, *Biochem. J.* **390**, 285 (2005).
23. M. Lasa-Benito, O. Marin, F. Meggio, L. A. Pinna, *FEBS Lett.* **382**, 149 (1996).
24. B. Barylko *et al.*, *J. Biol. Chem.* **276**, 7705 (2001).
25. F. F. Correia *et al.*, *J. Bacteriol.* **188**, 8360 (2006).
26. N. Kannan, S. S. Taylor, Y. Zhai, J. C. Venter, G. Manning, *PLoS Biol.* **5**, e17 (2007).
27. A. C. Groth, M. Fish, R. Nusse, M. P. Calos, *Genetics* **166**, 1775 (2004).
28. M. P. Zeidler, N. Perrimon, D. I. Strutt, *Dev. Biol.* **228**, 181 (2000).
29. We thank M. Simon for communicating unpublished observations; the Developmental Studies Hybridoma Bank and the Bloomington stock center for antibodies and *Drosophila* stocks; C. Rauskolb and T. Miller for comments on the manuscript; E. Cho and A. Xu for initial Fj, Fat, and Ds expression plasmids; and C. Rauskolb for assistance in generating images for figure panels. Supported by the Howard Hughes Medical Institute (K.D.I.), NIH grants CA123071, GM061126 (R.S.H.) and GM078620 (K.D.I.), and a fellowship from the Yamada Science Foundation (H.O.I.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/401/DC1
Materials and Methods
Figs. S1 to S8
References

21 March 2008; accepted 16 June 2008
10.1126/science.1158159

Signal-Mediated Dynamic Retention of Glycosyltransferases in the Golgi

Linna Tu, William C. S. Tai,* Lu Chen, David K. Banfield†

Golgi-resident glycosyltransferases are a family of enzymes that sequentially modify glycoproteins in a subcompartment-specific manner. These type II integral membrane proteins are characterized by a short cytoplasmically exposed amino-terminal tail and a luminal enzymatic domain. The cytoplasmic tails play a role in the localization of glycosyltransferases, and coat protein complex I (COPI) vesicle–mediated retrograde transport is also involved in their Golgi localization. However, the tails of these enzymes lack known COPI-binding motifs. Here, we found that Vps74p bound to a pentameric motif present in the cytoplasmic tails of the majority of yeast Golgi-localized glycosyltransferases, as well as to COPI. We propose that Vps74p maintains the steady-state localization of Golgi glycosyltransferases dynamically, by promoting their incorporation into COPI-coated vesicles.

The Golgi apparatus functions both as a sorting hub for newly synthesized proteins and as the location of posttranslational modification of glycoproteins. Protein glycosylation is initiated in the endoplasmic reticulum (ER). Glycoproteins then undergo extensive sequential carbohydrate chain elongations and elaborations mediated by a group of Golgi-localized glycosyltransferases. Although they catalyze distinct glycosylation reactions and display characteristic cisternal distributions, Golgi-resident glycosyltransferases are all type II integral membrane proteins consisting of a short cytoplasmically exposed N terminus (tail), a single membrane-spanning region, and a luminally orientated stalk region and catalytic domain.

A variety of features have been identified that influence the localization of these enzymes (*1–3*): the transmembrane domain (*4–8*), the luminally

---

Department of Biology, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, Special Administrative Region (SAR) of the People's Republic of China.

*Present address: Biomedical Science, School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Hong Kong, SAR of the People's Republic of China.

†To whom correspondence should be addressed. E-mail: bodkb@ust.hk

oriented noncatalytic region (*6*, *8*), interactions between catalytic domains (*9*, *10*), and the cytoplasmic tail (*11*–*13*). The steady-state distribution of these enzymes is maintained by a dynamic process that involves their retrieval from late Golgi cisterna to early cisterna, as well as between the Golgi and the ER, whereupon they transit back to the cisterna on which they function (*14*, *15*). Although the features of glycosyltransferases that direct a particular enzyme to one cisterna or another at steady-state varies, presumably coat protein complex I (COPI)–coated vesicles are the predominant means by which the majority of these enzymes are recycled (*14*, *16*). However, Golgi-resident glycosyltransferases lack canonical COPI-binding motifs in their cytoplasmic tails, and we sought to identify proteins that could bind the tails of these enzymes and could facilitate their incorporation into COPI-coated vesicles.

We identified Vps74p in a genetic screen for dosage suppressors of the lethality associated with deletion of the essential Golgi-resident SNARE (soluble *N*-ethylmaleimide–sensitive factor attachment protein receptor), Sft1p, a protein required for retrograde traffic within the Golgi (*17*). Vps74p is a member of a protein family that includes the mammalian proteins GPP34 (GmX33α) and GPP34R (GmX33β) (*18*–*21*). Although deletion of *VPS74* has no obvious effects on yeast cell growth, *vps74*Δ cells are known to missort carboxypeptidase Y (CPY), which suggests that Vps74p is involved in protein trafficking (*22*). We observed that *vps74*Δ cells were defective in the processing of the glycosylphosphatidylinositol (GPI)–anchored protein Gas1p. Gas1p is modified by the addition of N- and O-linked carbohydrates as it traverses the Golgi en route to the cell surface (Fig. 1A) (*23*). Comparisons of the fate of Gas1p in yeast strains lacking enzymes involved in N- and O-glycosylation revealed a similar processing defect in cells lacking the α1,2-mannosyltransferase, Kre2p (Fig. 1, A and B) (*24*). We considered the possibility that Vps74p might be required for Gas1p transport and examined the trafficking of a functional Gas1p–monomeric red fluorescent protein (Gas1p-mRFP) fusion protein (*25*) in cells lacking Vps74p. However, the location of Gas1p-mRFP in *vps74*Δ cells was indistinguishable from that in wild-type cells (Fig. 1C).

Like GPP34 (*18*), Vps74p is a peripheral membrane protein (Fig. 1, D and E) and, as such, would only be capable of direct physical interaction with the cytoplasmically orientated N-terminal 11 amino acids of Kre2p. We constructed a protein hybrid in which the cytoplasmic and transmembrane regions of Kre2p were fused to the N-terminal end of mRFP (Kre2CT-mRFP). In wild-type cells, Kre2CT-mRFP was localized to numerous puncta throughout the cytoplasm, which partially overlapped that of a green fluorescent protein (GFP) fusion to the Golgi-resident protein, Rer1p (GFP-Rer1p) (Fig. 1F). Furthermore, as do Rer1p (*26*) and the Mnn9p glycosyltransferase complex (*16*), Kre2CT-GFP cycled between the Golgi and the ER via a COPI- and SNARE–dependent mechanism (fig. S1). In *vps74*Δ cells, however, Kre2CT-mRFP was localized in the lumen of the vacuole (Fig. 1F). This effect was not due to a general deficiency in the retention of Golgi-resident membrane proteins, because the localization of GFP-Rer1p to the Golgi was unaffected. These findings are consistent with a role for Vps74p in the dynamic retention of Kre2p in the Golgi and identify the N-terminal 11 amino acids of the protein as sufficient to mediate retention.

Yeast cells in which particular genes involved in glycosylation have been deleted show reduced viability in the presence of calcofluor white (CW). Cells lacking *VPS74* were more sensitive to CW than *kre2*Δ cells (fig. S2), so we reasoned that Vps74p might mediate the Golgi localization of additional glycosyltransferases. When *VPS74* was deleted from strains expressing Mnn9-GFP, Mnn2-GFP, or Ktr6CT-GFP, these proteins were indeed mislocalized—either to a vesicle-like haze (Mnn9-GFP), or to the lumen of the vacuole (Mnn2-GFP and Ktr6CT-GFP) (Fig. 2A). Alignment of the N-terminal cytoplasmic tails of members of the yeast glycosyltransferase family revealed the presence of a semiconserved motif (F/L)-(L/V)-(S/T) (*27*) in the cytoplasmic tails of several members of the family (Table 1). We substituted Ala for Phe^4-Leu^5-Ser^6 (FLS) in Kre2p and assessed the fate of these substitutions on the localization of kre2 (AAA^6)CT-GFP, as well as on the capacity of the mutant full-length protein, kre2(AAA^6), to process Gas1p in *kre2*Δ cells. kre2(AAA^6)CT-GFP was mislocalized to the interior of the vacuole in wild-type cells (Fig. 2B), and cells expressing kre2(AAA^6) as their sole source of the enzyme were indistinguishable from cells lacking the *KRE2* gene in their inability to fully process Gas1p (Fig. 2C). We suggest that the defect in Gas1p processing was related to the instability of

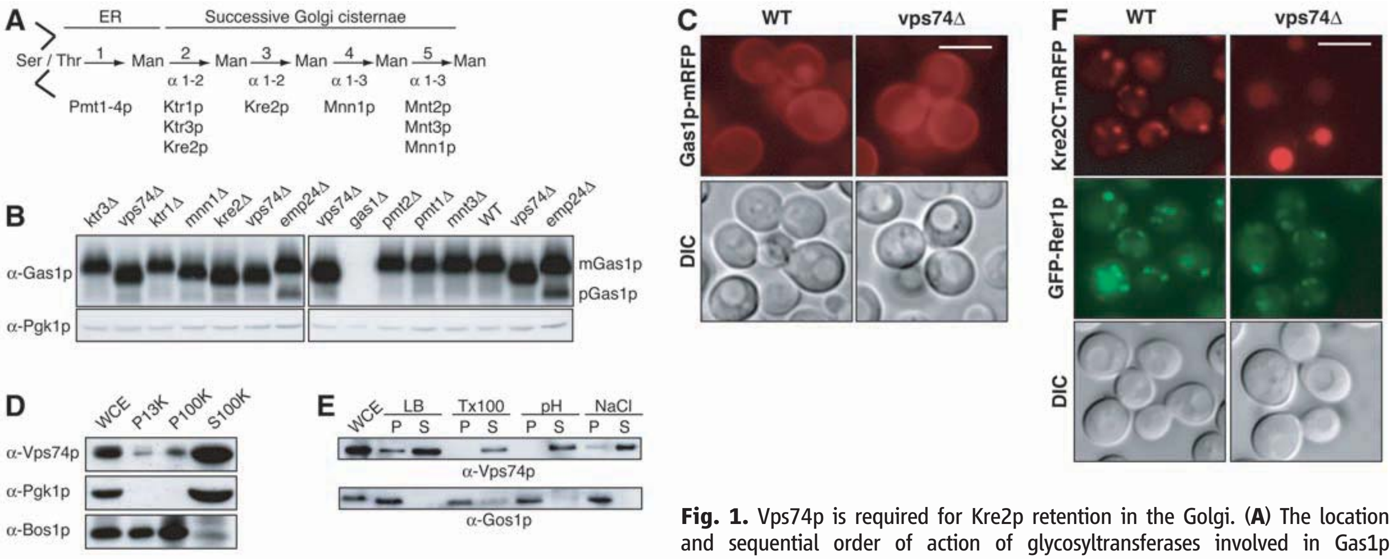


Fig. 1. Vps74p is required for Kre2p retention in the Golgi. (**A**) The location and sequential order of action of glycosyltransferases involved in Gas1p processing (*29*). (**B**) *vps74*Δ cells have a flaw similar to the Gas1p-processing defect of cells lacking Kre2p. The relative positions of fully processed and glycosylated Gas1p (mGas1p) and the ER-resident precursor form of Gas1p (pGas1p) are indicated. (**C**) Transport of Gas1p-mRFP to the cell surface is unaffected in *vps74*Δ cells. Scale bar, 5 μm. (**D**) Vps74p binds to membranes. Whole-cell extracts (WCEs) were subjected to sequential rounds of centrifugation at 13,000*g* (13K) and 100,000*g* (100K), and the pellet (P) and supernatants (S) were subjected to SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and immunostaining. (**E**) Vps74p is a peripheral membrane protein. Yeast WCEs were extracted with lysis buffer alone (LB) or with 1% Triton X100 (Tx100), 0.1 M Na_2CO_3 pH 11.5 (pH), or 1 M NaCl (NaCl) in lysis buffer. (**F**) Kre2CT-mRFP but not GFP-Rer1p is mislocalized to the vacuole lumen in *vps74*Δ cells. For a detailed description of reagents and methods used, see (*24*). DIC, differential interference contrast.

Kre2p in cells lacking Vps74p. Indeed, Kre2p was significantly less abundant in *vps74*Δ cells, and the abundance of kre2(AAA^6) in wild-type cells was very similar to that of Kre2p in *vps74*Δ cells (Fig. 2D). In addition, Gas1p processing in *kre2*Δ cells expressing N-terminal Kre2p hybrid proteins generated from glycosyltransferases containing FLS-like motifs [with the exception of Anp1 (Table 1)] was indistinguishable from that in wild-type cells (Fig. 3A). Further evidence that the FLS-like motif was important for recognition by Vps74p was obtained from a yeast two-hybrid assay in which the wild-type or FLS Ala-substituted cytoplasmic tail of Kre2p [kre2C3(AAA^6)] was used to assess binding by Vps74p. Although Vps74p was capable of binding to the wild-type cytoplasmic tail sequence of Kre2p, no interaction was observed between Vps74p and kre2C3(AAA^6) (Fig. 3B). Vps74p binds directly to the cytoplasmic tail of Kre2p, and substitutions to the FLS motif reduce binding (Fig. 3C and fig. S3).

The inability of the Anp1-Kre2 hybrid protein to restore Gas1p processing to *kre2*Δ cells (Fig. 3A) was puzzling, so we considered the possibility that the Vps74p glycosyltransferase recognition site contained residues in addition to the FLS-like motif. A recent study suggests an important role for dibasic sequences in the recycling of Mnn9p from the Golgi to the ER (*13*), and our yeast two-hybrid assays, Kre2p stability studies, and live-cell imaging experiments supported a requirement for the KR^8 peptide in the Vps74-dependent retention of Kre2p (fig. S4). In addition, we observed no requirement for the polar residue within the FLS-like motif of Kre2p, whereas F^4 and L^5 were essential for binding to Vps74p (fig. S5). Thus, the Vps74p-binding site on yeast Golgi-resident glycosyltransferases conforms to a consensus sequence of (F/L)-(L/I/V)-X-X-(R/K), a motif that is present in 16 Golgi-resident enzymes (Table 1). To test the validity of this proposal, we examined the fate of Mnn5-GFP, which contains a sequence that conforms to the consensus (LIXXK^6), in *vps74*Δ cells. Consistent with our prediction, Mnn5-GFP was mislocalized to the vacuole lumen in these cells (fig. S6).

Vps74p could facilitate Golgi retention of glycosyltransferases either by blocking anterograde trafficking of these enzymes from the Golgi or by assisting their recycling through Golgi cisterna or from the Golgi to the ER, in a process mediated via the COPI coat (fig. S1). To distinguish between these possibilities, we examined the fate of Kre2CT-GFP in cells that constitutively overexpressed Vps74p. We reasoned that if Vps74p functioned in an ER recycling pathway, overexpression of the protein might be expected to shift the steady-state distribution of Kre2CT-GFP from the Golgi to the ER. This is indeed the case; the effect of *VPS74* overexpression on the steady-state localization of Kre2CT-GFP was not due to an indirect block in ER export or to a general acceleration of retrograde transport (Fig. 3D).

A role for Vps74p in retrograde transport of Kre2CT-GFP was also supported by our genetic studies. In addition to *VPS74*, we also identified genes encoding components of the Golgi vesicle coat complex COPI or its substrates (fig. S7). Because Sft1p is required for retrograde transport within the Golgi (*17*), we surmised that overexpression of these genes compensated for this deficiency in *sft1*Δ cells. In addition, genetic interactions were observed between *VPS74* and mutants defective in ER-Golgi trafficking (fig. S7).

Given that the Mnn9p complex (*16*) and Kre2CT-GFP (fig. S1) cycle between the Golgi and ER in a COPI-dependent manner, we considered the possibility that Vps74p might function together with COPI to mediate the recycling

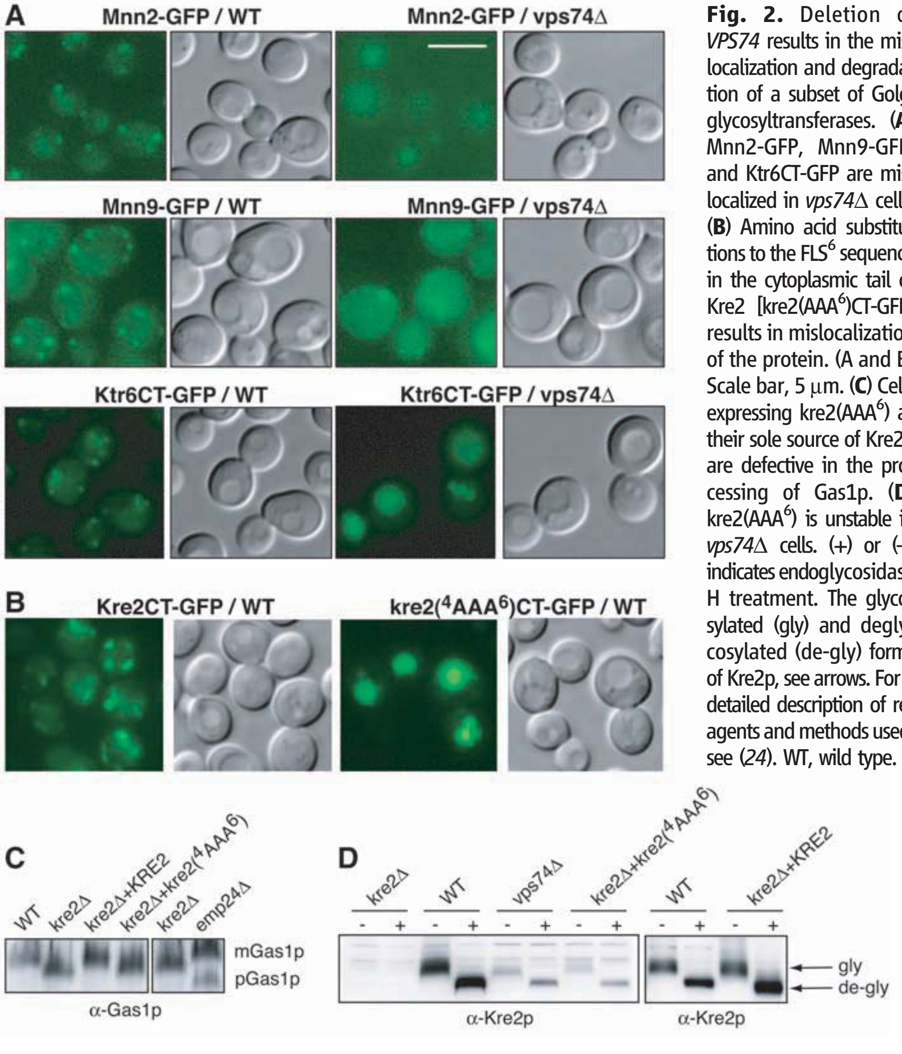


Fig. 2. Deletion of *VPS74* results in the mislocalization and degradation of a subset of Golgi glycosyltransferases. (**A**) Mnn2-GFP, Mnn9-GFP, and Ktr6CT-GFP are mislocalized in *vps74*Δ cells. (**B**) Amino acid substitutions to the FLS^6 sequence in the cytoplasmic tail of Kre2 [kre2(AAA^6)CT-GFP] results in mislocalization of the protein. (A and B) Scale bar, 5 μm. (**C**) Cells expressing kre2(AAA^6) as their sole source of Kre2p are defective in the processing of Gas1p. (**D**) kre2(AAA^6) is unstable in *vps74*Δ cells. (+) or (–) indicates endoglycosidase H treatment. The glycosylated (gly) and deglycosylated (de-gly) forms of Kre2p, see arrows. For a detailed description of reagents and methods used, see (*24*). WT, wild type.

Table 1. The amino acid sequences of the cytoplasmic tails of yeast Golgi-resident glycosyltransferases (*27*). FLS-like motifs are highlighted in bold italic font. Glycosyltransferases bearing presumptive Vps74p-binding (F/L)(L/I/V)-X-X-(R/K) motifs are highlighted in bold font and their corresponding motifs are underlined and in italics.

Enzyme	Amino acid sequence of the cytoplasmic tail
Kre2	MAL***FLS****KR*LLR
Ktr1	MAKIMIPASKQPVYKK
Ktr2	MQICKV***FLT****QVK*K
Ktr3	MSVHHKKKLMPKSA*LLIRK*YQKGIR
Ktr4	MR***FLS****KR*ILK
Ktr5	M*LLIRR*TINAFLGCIH
Ktr6	MHV***LLS****KK*IAR
Ktr7	MAIRLNPKVRR*FLLDK*CRQKR
Yur1	MAK
Mnt2	MRRKNR
Mnt3	MLKSLKSRR*LILKR*
Mnt4	M*VLRIR*RIKK
Mnn1	MLALRRF*ILNQR*SLR
Mnn2	M***LLT****KR*FSKLFK
Mnn4	MLQRISSKLHRR***FLS***G*LLRVK*HYPLRR
Mnn5	M*LIRLK*KRK
Mnn9	MSLS***LVS****YR*LRK
Anp1	MKYNNRKLSFNPTTVSIAGT**LLT**VF**FLT**R
Hoc1	MAKTTKRASSFRR
Gnt1	MR***LIS****KR*RIR
Och1	MSRKLSH*LIATR*KSK

of glycosyltransferases. Consistent with this hypothesis, Sec26p (βCOP) was dependent on Vps74p for its ability to support the growth of *sft1*Δ cells (fig. S8A), and Vps74p bound to coatomer (Fig. 3E), as well as to the trunk region of Sec26p and Ret2p (δCOP) in in vitro mixing assays (fig. S8B and Fig. 3F).

We suggest that Vps74p functions as a component of the Golgi-ER vesicular transport machinery as a glycosyltransferase-sorting receptor for the COPI coat and, in so doing, controls the steady-state distribution of these enzymes by preventing their transport beyond the Golgi. Additional support for this model comes from a recent study in which the requirement for Vps74p in the retention of glycosyltransferases was also reported (*28*). Although the human homologs of Vps74p, GPP34, and GPP34R suppress the phenotypes of *vps74*Δ cells (Fig. 3G and fig. S9), plants and protozoa appear to lack orthologs of Vps74p. Thus, multiple mechanisms (*1*–*3*) likely function to ensure the correct localization of Golgi-resident glycosyltransferases in cells.

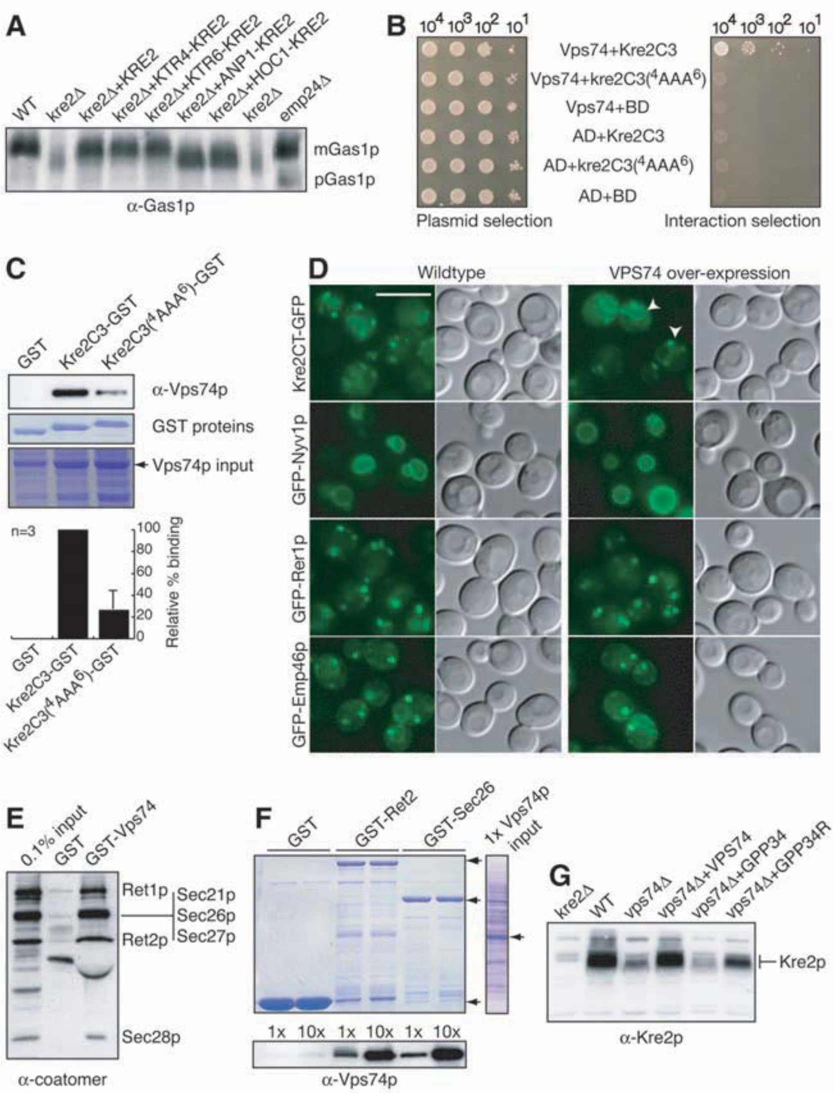


Fig. 3. Vps74p binds to the cytoplasmic tails of Golgi glycosyltransferases and to coatomer. (**A**) Glycosyltransferase-Kre2p chimeras bearing FLS-like motifs restore Gas1p processing in *kre2*Δ cells. (**B**) The FLS^6 sequence in the cytoplasmic tail of Kre2p is required for interaction with Vps74p in a yeast two-hybrid assay. AD and BD refer to the activation and binding domains of GAL4, respectively. (**C**) Vps74p binds directly to the cytoplasmic tail of Kre2p. (Top) Immunoblot of 5% copurified proteins; (middle and bottom) 20% of input proteins. (Graph) Quantification of the results (means ± SD). A fusion protein of kre2C3(AAA^6) with glutathione *S*-transferase [kre2C3(AAA^6)-GST] (26.4 ± 18.6%) and GST alone (0.1 ± 0.3%). (**D**) Overexpression of *VPS74* redistributes Kre2CT-GFP from the Golgi to the ER. The arrows highlight the ER localization of Kre2CT-GFP. Scale bar, 5 μm. (**E**) GST-Vps74p binds to coatomer from yeast WCEs. The binding of GST-Vps74p to coatomer was 7.5 ± 0.4 (means ± SD) times as great as that of GST alone. (**F**) Vps74p binds to full-length GST-Ret2p and to the trunk domain (amino acids 1 to 372) of GST-Sec26p. (Top, left) Proteins purified on glutathione beads after incubation with histidine $(\mathrm{His})_6$-tagged Vps74p; (top, right) 1× input. The arrows indicate the positions of the recombinant proteins. (Bottom) Immunostaining of Vps74p bound to GST fusion proteins. The relative amounts of Vps74p input proteins mixed with GST fusion proteins are shown at 1× and 10×. (**G**) GPP34R suppresses the instability of Kre2p in *vps74*Δ cells. For a detailed description of reagents and methods used, see (*24*).

References and Notes

1. K. J. Colley, *Glycobiology* **7**, 1 (1997).
2. A. S. Opat, C. van Vliet, P. A. Gleeson, *Biochimie* **83**, 763 (2001).
3. S. R. Pfeffer, *Annu. Rev. Biochem.* **76**, 629 (2007).
4. S. Munro, *EMBO J.* **10**, 3577 (1991).
5. J. Burke, J. M. Pettitt, H. Schachter, M. Sarkar, P. A. Gleeson, *J. Biol. Chem.* **267**, 24433 (1992).
6. T. R. Graham, V. A. Krasnov, *Mol. Biol. Cell* **6**, 809 (1995).
7. S. Munro, *EMBO J.* **14**, 4695 (1995).
8. T. Nilsson, C. Rabouille, N. Hui, R. Watson, G. Warren, *J. Cell Sci.* **109**, 1975 (1996).
9. T. Nilsson, P. Slusarewicz, M. H. Hoe, G. Warren, *FEBS Lett.* **330**, 1 (1993).
10. T. Nilsson *et al.*, *EMBO J.* **13**, 562 (1994).
11. M. Lussier, A. M. Sdicu, T. Ketela, H. Bussey, *J. Cell Biol.* **131**, 913 (1995).
12. A. S. Uliana, C. G. Giraudo, H. J. Maccioni, *Traffic* **7**, 604 (2006).
13. M. Okamoto, T. Yoko-o, T. Miyakawa, Y. Jigami, *Eukaryot. Cell* **7**, 310 (2008).
14. C. Rabouille, J. Klumperman, *Nat. Rev. Mol. Cell Biol.* **6**, 812 (2005).
15. M. A. Puthenveedu, A. D. Linstedt, *Curr. Opin. Cell Biol.* **17**, 369 (2005).
16. Z. Todorow, A. Spang, E. Carmack, J. Yates, R. Schekman, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13643 (2000).
17. D. K. Banfield, M. J. Lewis, H. R. Pelham, *Nature* **375**, 806 (1995).
18. A. W. Bell *et al.*, *J. Biol. Chem.* **276**, 5152 (2001).
19. C. C. Wu *et al.*, *Traffic* **1**, 963 (2000).
20. C. M. Snyder, G. A. Mardones, M. S. Ladinsky, K. E. Howell, *Mol. Biol. Cell* **17**, 511 (2006).
21. M. Nagano-Ito, S. Yoshikawa, M. Tamura, M. Tsurumaki, S. Ichikawa, *Gene* **400**, 82 (2007).
22. C. J. Bonangelino, E. M. Chavez, J. S. Bonifacino, *Mol. Biol. Cell* **13**, 2486 (2002).
23. L. Popolo, M. Vai, *Biochim. Biophys. Acta* **1426**, 385 (1999).
24. Materials and methods are available as supporting materials on *Science* Online.
25. M. Fujita, T. Yoko-o, Y. Jigami, *Mol. Biol. Cell* **17**, 834 (2006).
26. K. Sato, M. Sato, A. Nakano, *J. Cell Biol.* **152**, 935 (2001).
27. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr, and X, any amino acid.
28. K. R. Schmitz *et al.*, *Dev. Cell* **14**, 523 (2008).
29. P. A. Romero *et al.*, *Glycobiology* **9**, 1045 (1999).
30. We thank J. Wong and S. Ho for critically reading the manuscript and H. Reizman, A. Spang and A. Nakano for antibodies, plasmids, and strains. This work was supported by grants HKUST/02M, HKUST/05M and 660007 to D.K.B. from the Research Grants Council of Hong Kong.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/404/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 and S2
References

21 April 2008; accepted 20 June 2008
10.1126/science.1159411

Anomalous Type 17 Response to Viral Infection by CD8$^+$ T Cells Lacking T-bet and Eomesodermin

Andrew M. Intlekofer,[1,2] Arnob Banerjee,[1,2] Naofumi Takemoto,[1,2] Scott M. Gordon,[1,2] Caitlin S. DeJong,[1,2] Haina Shin,[5] Christopher A. Hunter,[4] E. John Wherry,[5] Tullia Lindsten,[1,3] Steven L. Reiner[1,2]*

When intracellular pathogens invade mammalian hosts, naïve CD8$^+$ T cells differentiate into cytotoxic killers, which lyse infected target cells and secrete cytokines that activate intracellular microbicides. We show that CD8$^+$ T cells deficient in the transcription factors T-bet and eomesodermin (Eomes) fail to differentiate into functional killers required for defense against lymphocytic choriomeningitis virus. Instead, virus-specific CD8$^+$ T cells lacking both T-bet and Eomes differentiate into an interleukin-17–secreting lineage, reminiscent of the helper T cell fate that has been implicated in autoimmunity and extracellular microbial defense. Upon viral infection, mice with T cells lacking both T-bet and Eomes develop a CD8$^+$ T cell–dependent, progressive inflammatory and wasting syndrome characterized by multi-organ infiltration of neutrophils. T-bet and Eomes, thus, ensure that CD8$^+$ T cells adopt an appropriate course of intracellular rather than extracellular destruction.

The T-box transcription factors T-bet and Eomes function in a variety of immune cells to regulate the expression of genes involved in Type 1 immunity (*1*, *2*). In CD4$^+$ T cells, T-bet plays a critical role in driving induction of interferon-γ (IFN-γ) and other aspects of the Type 1 genetic program (*3*–*7*). T-bet also functions as a critical and nonredundant regulator of the development and function of natural killer (NK) T cells (*8*). In contrast, both T-bet and Eomes are expressed in CD8$^+$ T cells and NK cells, suggesting that they might function redundantly to regulate the cytotoxic effector function of these cells (*8*–*10*). Although studies of T-bet–deficient mice have helped elucidate numerous functions of T-bet, the precise role of Eomes in effector and memory CD8$^+$ T cell differentiation has yet to be resolved because unconditional genetic ablation of *Eomes* results in early embryonic lethality (*11*).

To examine the function of Eomes in T cells, we used homologous recombination to generate mice carrying a *lox*P-flanked allele of *Eomes* (*12*). We bred this strain to mice expressing Cre recombinase under control of CD4 promoter/enhancer elements, thereby allowing selective deletion of *Eomes* in both CD4$^+$ and CD8$^+$ T cells. Mice with homozygous T cell–specific deletion of *Eomes* (here referred to as "Eomes KO") were, in turn, crossed to mice with germline deletion of *Tbx21* ("T-bet KO"), creating mice with combined deficiency for both T-bet and Eomes in the T cell compartment ("Double KO"). T-bet KO, Eomes KO, and Double KO mice exhibited no gross developmental abnormalities or disease phenotypes. Conventional CD4$^+$ and CD8$^+$ T cell development in T-bet KO, Eomes KO, and Double KO mice was unaffected, with numbers and ratios of thymocyte subsets that were similar to those of wild-type mice (*12*). We observed relatively normal CD4$^+$ to CD8$^+$ T cell ratios in the spleen and lymph nodes of T-bet KO, Eomes KO, and Double KO mice (*12*), although Eomes KO and Double KO mice exhibited a deficiency in memory-phenotype (CD44^{hi}CD122^{hi}) CD8$^+$ T cells, as predicted by previous findings (*9*).

Infection with lymphocytic choriomeningitis virus (LCMV) induces robust proliferation and cytotoxic differentiation of CD8$^+$ T cells, which are essential to eliminate virus from the host (*13*). To test the functional capacity of CD8$^+$ T cells deficient for T-bet and/or Eomes, we infected wild-type, T-bet KO, Eomes KO, and Double KO mice with LCMV and assessed their ability to control the viral infection. Virus was efficiently cleared from the spleens and kidneys of wild-type, T-bet KO, and Eomes KO mice (Fig. 1A). In contrast, Double KO mice harbored high titers of persistent virus, comparable to lymphocyte-deficient *Rag1*$^{-/-}$ ("Rag KO") mice (Fig. 1A).

Consistent with the inability of Double KO mice to control viral replication, LCMV-specific CD8$^+$ T cells from Double KO mice exhibited a substantial defect in IFN-γ expression (Fig. 1B) and target cell cytotoxicity (Fig. 1C) in response to stimulation with viral peptides. Double KO cells also exhibited marked impairment in the

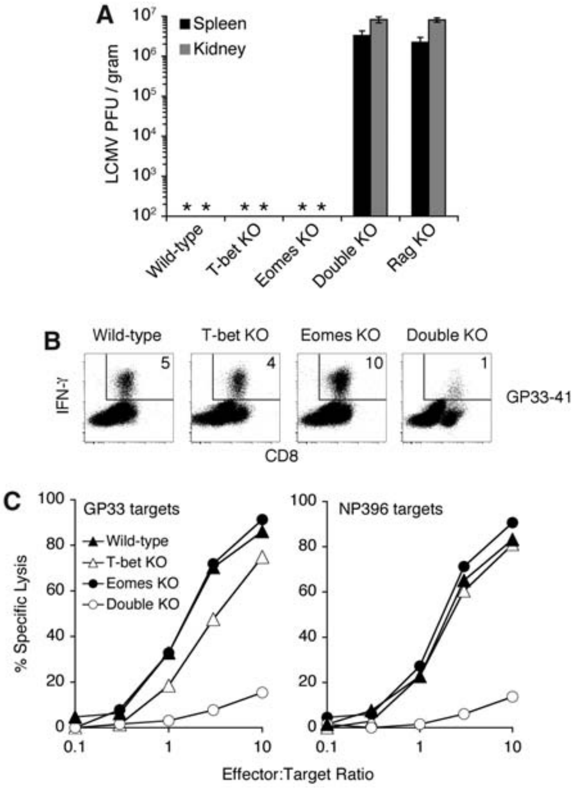


Fig. 1. Mice with T cells lacking both T-bet and Eomes fail to control LCMV infection. (**A**) Quantification of LCMV virus in spleen and kidney from mice of the indicated genotype 23 days after infection. Detection threshold of limiting dilution assay is ~2 × 10^2 plaque-forming units per gram (PFU/g); an asterisk (*) indicates that virus was undetectable. Values represent the mean ± SEM of at least three mice per genotype. (**B** and **C**) Defective IFN-γ expression and cytolysis of target cells by CD8$^+$ T cells lacking T-bet and Eomes. Splenocytes from mice of the indicated genotypes were isolated 8 days after infection with LCMV. (B) Flow cytometry of IFN-γ production by CD8$^+$ T cells in response to stimulation with LCMV-derived peptide GP33-41. Percentage of live cells expressing IFN-γ is indicated. (C) Cytolytic activity of virus-specific CD8$^+$ T cells. Equalized numbers of GP33- or NP396-specific CD8$^+$ T cells were added to target cells pulsed with LCMV-derived peptides GP33-41 or NP396-404. Percent specific lysis was determined after 24 hours (*12*). Virus-specific CD8$^+$ T cells from Double KO mice also exhibited impaired accumulation and abnormal expression of numerous surface receptors, reflecting both persistent antigen exposure and abnormal differentiation (*12*). Results are representative of at least three independent experiments.

---

[1] Abramson Family Cancer Research Institute, University of Pennsylvania, Philadelphia, PA 19104, USA. [2] Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [3] Department of Pathology, University of Pennsylvania, Philadelphia, PA 19104, USA. [4] Department of Pathobiology, University of Pennsylvania, Philadelphia, PA 19104, USA. [5] Immunology Program, The Wistar Institute, Philadelphia, PA 19104, USA.

*To whom correspondence should be addressed at the Abramson Family Cancer Research Institute, University of Pennsylvania, Building BRB II/III, Room 414, 421 Curie Boulevard, Philadelphia, PA 19104–6160, USA. E-mail: sreiner@mail.med.upenn.edu

induction of granzyme B protein and perforin mRNA (*12*). In addition to impaired cytotoxic differentiation, virus-specific Double KO CD8$^+$ T cells exhibited defective accumulation (reduced by a factor of 4.5), likely owing to impaired proliferation and survival (*12*). Although the reduced frequency of antigen-specific cells may contribute to impaired clearance of virus, the severe defect in cytotoxic gene expression (*12*) and function (Fig. 1C) of Double KO cells is wholly evident when cellularity is normalized. Antiviral CD8$^+$ T cell effector differentiation can, thus, withstand the loss of either T-bet or Eomes, but the loss of both factors is incompatible with appropriate antiviral host defense.

During the first week of LCMV infection, there were no apparent differences in the health of mice of all four genotypes. At approximately 1 week after infection, however, Double KO mice began to lose weight, which steadily progressed over the course of several weeks (Fig. 2A) and invariably necessitated euthanasia. In addition to the virus-induced weight loss observed in Double KO mice, we found evidence of extensive organ pathology. Examination of the spleens of LCMV-infected Double KO mice, revealed an excessive neutrophil response (Fig. 2B), a pattern typically associated with bacterial or fungal rather than viral infection (*14*–*16*). Histologic examination revealed widespread immunopathology, characterized by heavy neutrophil infiltration and architectural destruction of numerous organs (Fig. 2, C and D). Immunodeficient Rag KO mice did not lose weight (Fig. 2A) nor develop destructive organ pathology despite comparable viral burdens (Fig. 1A), suggesting that the immunopathology of Double KO mice was lymphocyte-dependent.

Interleukin-17 (IL-17)–secreting CD4$^+$ helper T cells are thought to orchestrate beneficial and pathological immune responses by the recruitment and activation of neutrophils (*14*–*19*). The presence of widespread neutrophilic pathology in Double KO mice prompted us to reevaluate the lineage markers of virus-specific CD8$^+$ T cells. The cytokine profile of CD8$^+$ T cells from Double KO mice revealed substantial expression of IL-17 in response to stimulation with virus-derived peptides (Fig. 3A), with little evidence of IFN-γ (Fig. 3A), IL-4, IL-10, or Foxp3 expression (Fig. 3, B and C), indicating that virus-specific CD8$^+$ T cells from Double KO mice were undergoing skewed differentiation exclusively toward the Type 17 fate. Indeed, the transcriptional profile of virus-specific CD8$^+$ T cells from Double KO mice was consistent with bona fide Type 17 differentiation (Fig. 3D), with substantial induction of the lineage-specifying transcription factor (RORγt), signature cytokines (IL-17A, IL-21, and IL-22), and characteristic cytokine receptor (IL-23R) of the Type 17 fate (*20*–*24*). Although virus-specific CD8$^+$ T cells from wild-type and Eomes KO mice did not express IL-17, we found some evidence of derepressed Type 17 markers in the CD8$^+$ T cells from T-bet KO mice (Fig. 3, A and D). In separate experiments, we determined that IL-17 expression by CD8$^+$ T cells was not contingent on T-bet deficiency in CD4$^+$ T cells or cells of the innate immune system (*12*). It is yet unclear whether T-bet KO mice are spared from immunopathology because of the decreased intensity of the Type 17 response, the lack of persistent antigen stimulation, or a combination of these factors.

To determine whether the anomalous T cell response was required for immunopathology observed in Double KO mice, we performed antibody-mediated depletion of either CD4$^+$ or CD8$^+$ T cells during LCMV infection. Despite the finding that CD4$^+$ T cells from Double KO mice also express IL-17 (*12*), depletion of CD4$^+$ T cells did not alter the course of the wasting disease (Fig. 4A) or ameliorate the pathological findings. Depletion of CD8$^+$ T cells, however, prevented the virus-induced wasting syndrome (Fig. 4A), blood neutrophilia (Fig. 4B), and organ pathology (Fig. 4C), demonstrating that the anomalous CD8$^+$ T cell subset is indeed required to

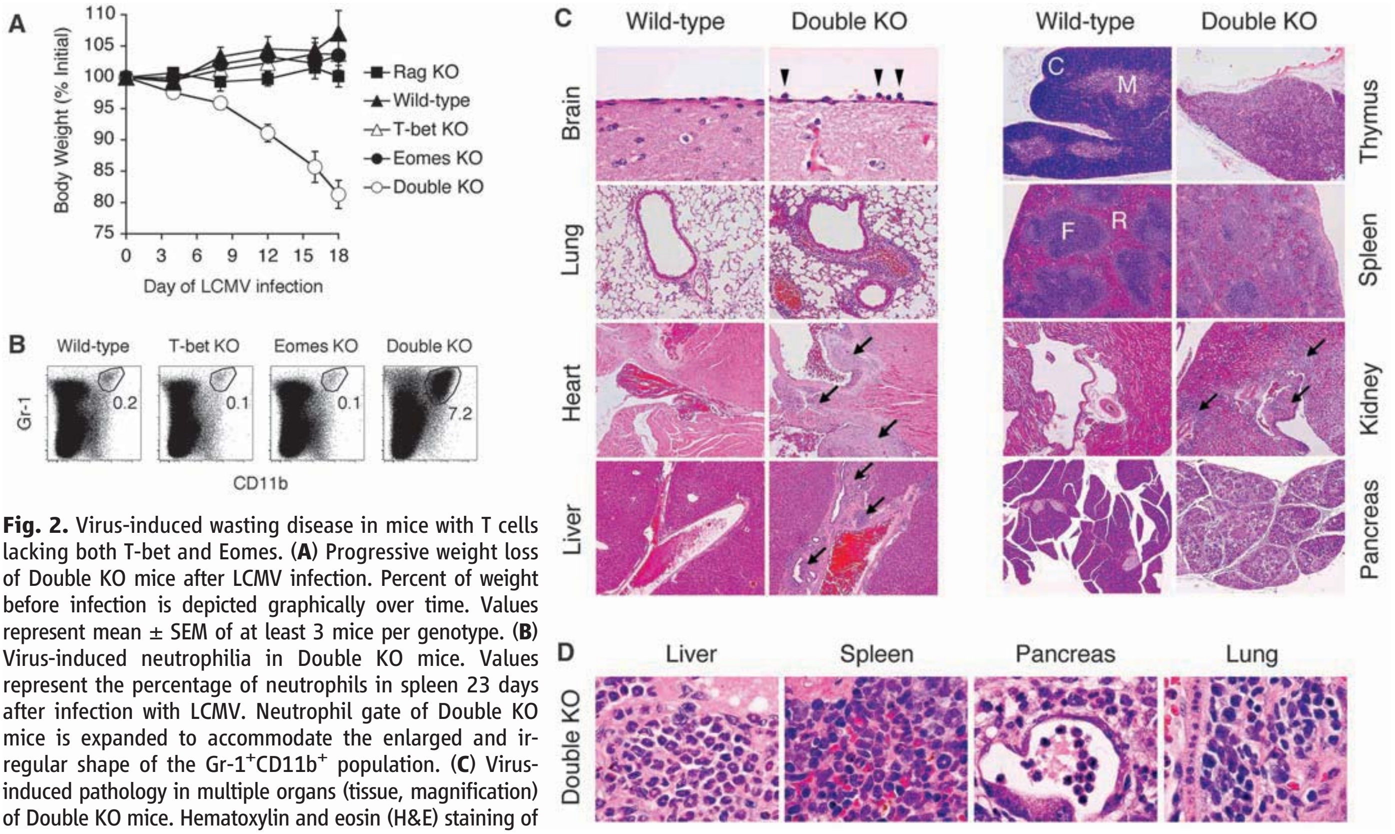


Fig. 2. Virus-induced wasting disease in mice with T cells lacking both T-bet and Eomes. (**A**) Progressive weight loss of Double KO mice after LCMV infection. Percent of weight before infection is depicted graphically over time. Values represent mean ± SEM of at least 3 mice per genotype. (**B**) Virus-induced neutrophilia in Double KO mice. Values represent the percentage of neutrophils in spleen 23 days after infection with LCMV. Neutrophil gate of Double KO mice is expanded to accommodate the enlarged and irregular shape of the Gr-1$^+$CD11b$^+$ population. (**C**) Virus-induced pathology in multiple organs (tissue, magnification) of Double KO mice. Hematoxylin and eosin (H&E) staining of brain (meninges), 60×; lung (airway with associated vessel), 10×; heart (base), 5×; liver (portal vessel), 5×; thymus (C, cortex; M, medulla), 5×; spleen (F, follicle; R, red pulp), 5×; kidney (hilar vessel near calyx), 5×; pancreas, 5×. Arrowheads indicate neutrophils; arrows indicate immune cell infiltration. All organs are from 28 days after infection, except for kidney and pancreas (23 days). (**D**) High-power magnification (100×) demonstrating neutrophilic infiltration of liver, spleen, and pancreas, and mixed monocytic/neutrophilic infiltration of lung. Results are representative of two independent experiments with more than six mice per group.

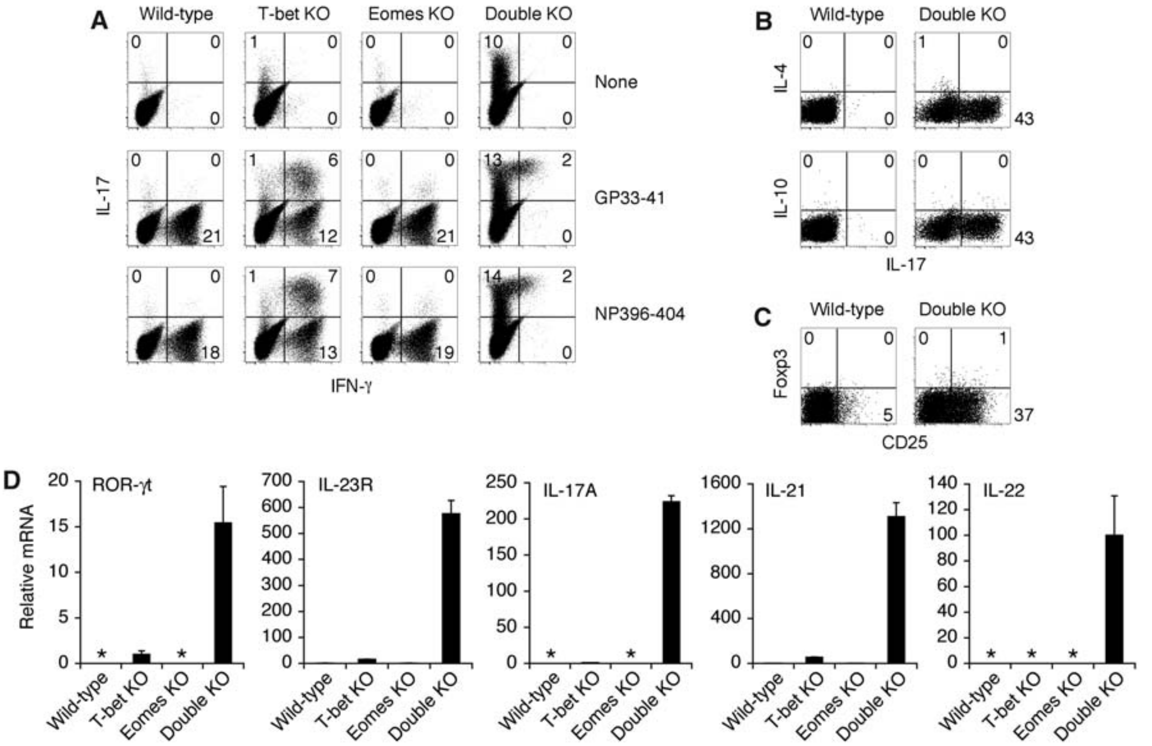


Fig. 3. Type 17 effector differentiation of CD8$^+$ T cells lacking both T-bet and Eomes. (**A**) IFN-γ and IL-17 production by LCMV-specific CD8$^+$ T cells from Double KO mice. Splenocytes from mice of the indicated genotypes were isolated 8 days after LCMV infection and given no stimulation (top row) or stimulated with LCMV-derived peptides GP33-41 (middle row) or NP396-404 (bottom row); cytokine production was assessed by intracellular staining; plots show CD8$^+$ events; numbers indicate the percentage of CD8$^+$ T cells expressing the indicated cytokine. Expression of IL-17 by Double KO CD8$^+$ T cells in the absence of restimulation (upper right plot) could be due to persistent virus in vivo (Fig. 1A) and/or constitutive expression of this cytokine (*12*). In addition to the abnormal CD8$^+$ T cell response, CD4$^+$ T cells from both T-bet KO and Double KO expressed IL-17 in response to stimulation with LCMV-derived peptide GP61-80 (*12*). Additional controls were performed to implicate the CD8$^+$ T cell–intrinsic loss of T-box factors as the cause of anomalous differentiation: Analysis of IFN-γ- and perforin-deficient mice suggested that IL-17 production by CD8$^+$ T cells from Double KO mice was not solely attributable to impaired cytotoxic effector differentiation (*12*). In separate experiments, we determined that IL-17 expression by CD8$^+$ T cells was not contingent on T-bet deficiency in CD4$^+$ T cells or cells of the innate immune system (*12*). (**B** and **C**) Lack of IL-4, IL-10, or Foxp3 expression by CD8$^+$ T cells from Double KO mice. Splenocytes from indicated genotypes were isolated 8 days after LCMV infection. Cells were either stimulated with antibody to CD3 (anti-CD3) (B) or examined directly (C). Cytokine production (B) or Foxp3 expression (C) were assessed by intracellular staining. Numbers indicate the percentage of CD8$^+$ T cells expressing the indicated cytokine or lineage marker. Despite their defective cytotoxocity and IFN-γ expression, virus-specific CD8$^+$ T cells from Double KO mice were able to produce tumor necrosis factor–α (TNF-α) and IL-2 (*12*). (**D**) Type 17 transcriptional profile of LCMV-specific CD8$^+$ T cells from Double KO mice. GP33-specific CD8$^+$ T cells were isolated by the use of peptide-MHC tetramers from indicated genotypes 8 days after LCMV infection. mRNA expression was assessed by quantitative real-time polymerase chain reaction. An asterisk (*) indicates undetectable signal. Lowest detectable sample was arbitrarily set at a value of 1, except for IL-22, where only Double KO had detectable signal. Values represent the mean ± SEM of triplicate determinations normalized to hypoxanthine guanine phosphoribosyl transferase. Results are representative of at least two independent experiments.

mediate the virus-induced wasting disease. Because CD8 is also expressed on a subset of dendritic cells, we cannot exclude the possibility that reversal of the lymphocyte-dependent wasting syndrome and immunopathology could be due to the combined loss of Double KO CD8$^+$ T cells and CD8$^+$ DCs. Despite the defective accumulation of antigen-specific CD8$^+$ T cells, the virus-induced pathology of Double KO mice is distinct from the functional ignorance of virus and the absence of pathology in Rag KO mice.

When a pathogen invades a mammalian host, CD4$^+$ T cells differentiate into various fates, which mobilize the most appropriate intracellular or extracellular defense mechanisms befitting the nature of the pathogen. In contrast, CD8$^+$ T cells characteristically provide defense against intracellular infections by differentiating into cytotoxic killers. Based on the incomplete loss of effector function in T-bet KO CD8$^+$ T cells (*4*) and the surprisingly dense defect imposed on this lineage by a dominant-interfering form of T-bet, we proposed that a second T-box factor expressed in CD8$^+$ T cells, Eomes, might account for the T-bet–independent effector function of CD8$^+$ T cells (*10*). The present findings offer genetic evidence to suggest that T-bet and Eomes indeed redundantly activate a transcriptional network required for CD8$^+$ T cell participation in defense against intracellular pathogens, while simultaneously repressing the Type 17 genetic program involved in defense against extracellular bacteria and fungi.

It remains to be determined what nonoverlapping functions are mediated by these two T-box transcription factors. One provisional model suggests that they may play complementary roles in adjusting the balance between terminally differentiated effector cells and self-renewing memory CD8$^+$ T cells (*25*). Future studies will be directed toward understanding whether Type 17 CD8$^+$ T cells play a protective role in host defense against any classes of pathogens, or whether this represents a vestigial and pathological response for the CD8$^+$ T cell that has primarily evolved for intracellular defense.

References and Notes

1. L. H. Glimcher, M. J. Townsend, B. M. Sullivan, G. M. Lord, *Nat. Rev. Immunol.* **4**, 900 (2004).
2. L. H. Glimcher, *Nat. Immunol.* **8**, 448 (2007).
3. S. J. Szabo *et al.*, *Cell* **100**, 655 (2000).
4. S. J. Szabo *et al.*, *Science* **295**, 338 (2002).
5. A. C. Mullen *et al.*, *Science* **292**, 1907 (2001).
6. A. C. Mullen *et al.*, *Nat. Immunol.* **3**, 652 (2002).
7. M. Afkarian *et al.*, *Nat. Immunol.* **3**, 549 (2002).
8. M. J. Townsend *et al.*, *Immunity* **20**, 477 (2004).
9. A. M. Intlekofer *et al.*, *Nat. Immunol.* **6**, 1236 (2005).
10. E. L. Pearce *et al.*, *Science* **302**, 1041 (2003).
11. A. P. Russ *et al.*, *Nature* **404**, 95 (2000).
12. Supporting material is available on *Science* Online.
13. M. A. Williams, M. J. Bevan, *Annu. Rev. Immunol.* **25**, 171 (2007).
14. R. Medzhitov, *Nature* **449**, 819 (2007).
15. S. L. Reiner, *Cell* **129**, 33 (2007).
16. C. T. Weaver, R. D. Hatton, P. R. Mangan, L. E. Harrington, *Annu. Rev. Immunol.* **25**, 821 (2007).
17. H. Park *et al.*, *Nat. Immunol.* **6**, 1133 (2005).
18. L. E. Harrington *et al.*, *Nat. Immunol.* **6**, 1123 (2005).
19. P. R. Mangan *et al.*, *Nature* **441**, 231 (2006).
20. I. I. Ivanov *et al.*, *Cell* **126**, 1121 (2006).
21. L. Zhou *et al.*, *Nat. Immunol.* **8**, 967 (2007).
22. Y. Zheng *et al.*, *Nature* **445**, 648 (2007).
23. R. Nurieva *et al.*, *Nature* **448**, 480 (2007).
24. T. Korn *et al.*, *Nature* **448**, 484 (2007).

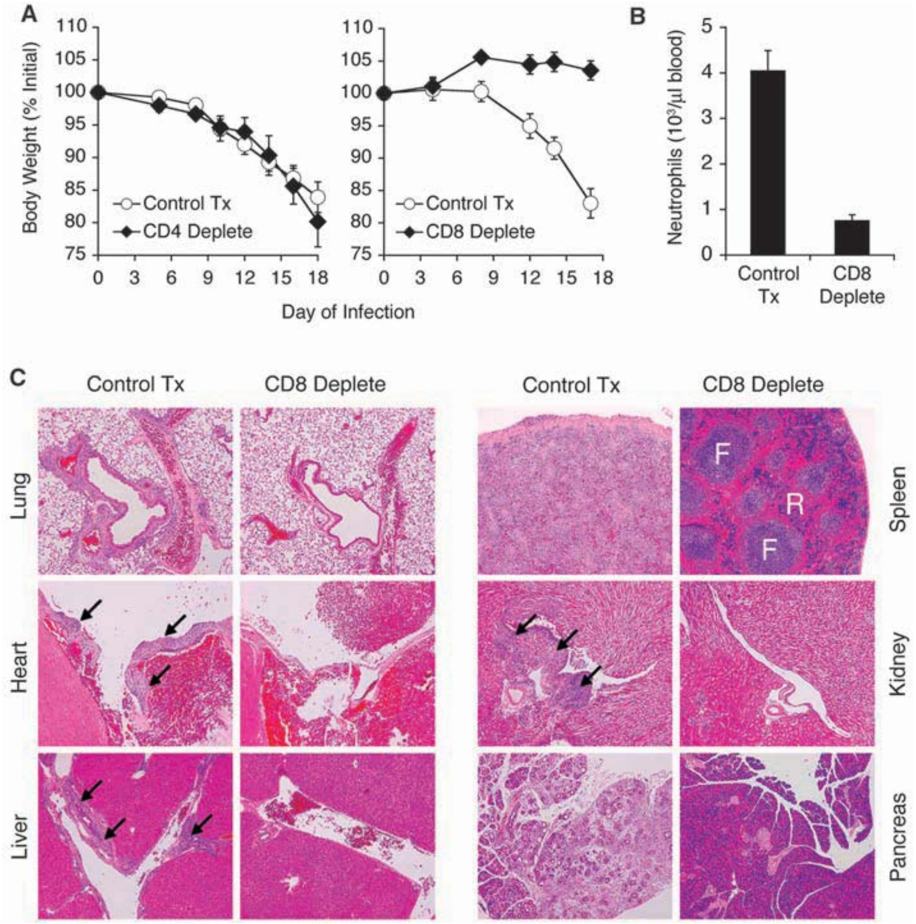


Fig. 4. CD8$^+$ T cell depletion prevents virus-induced wasting disease in Double KO mice. (**A**) Double KO mice were treated with phosphate-buffered saline ("Control Tx"), anti-CD4 ["CD4 Deplete," monoclonal antibody (mAb) GK1.5, 0.5 mg intraperitoneally (ip), left], or anti-CD8 ("CD8 Deplete," mAb 2.43, 0.5 mg ip, right) 1 day before, 1 day after, and every 3 days after LCMV infection. Weight at the indicated day after infection is displayed as a percentage of weight before infection. (**B**) Neutrophils were quantified in the blood of control or anti-CD8–treated mice at day 18 after infection by automated complete blood count with differential (right). Number of neutrophils per microliter of blood is depicted graphically. (**C**) H&E staining of organs from control-treated or CD8$^+$ T cell–depleted Double KO mice at day 18 after infection. Lung (airway with associated vessel), 10×; heart (base), 5×; liver (portal vessel), 5×; spleen (F, follicle; R, red pulp), 5×; kidney (hilar vessel near calyx), 5×; pancreas, 5×. Arrows indicate immune cell infiltration. Values for (A) and (B) represent the mean ± SEM of three individual mice per treatment group. Results are representative of two independent experiments.

25. A. M. Intlekofer *et al.*, *J. Exp. Med.* **204**, 2015 (2007).

26. We are grateful to E. Allenspach, S. Blackburn, J. Chang, D. Garalde-Intlekofer, I. Kinjyo, P. Mangan, V. Palanivel, F. Schambach, H. Yu, and Q. C. Yu for assistance and discussion. These studies were supported by National Institutes of Health (grants AI042370, AI061699, AI076458, AI071309, AI042334, and AI007532) and the Abramson Family Cancer Research Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/408/DC1
Materials and Methods
Figs. S1 to S9
References

29 April 2008; accepted 11 June 2008
10.1126/science.1159806

Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP

N. Sudarsan,[1] E. R. Lee,[2] Z. Weinberg,[2] R. H. Moy,[3] J. N. Kim,[2] K. H. Link,[1] R. R. Breaker[1,2,3]*

Cyclic di-guanosine monophosphate (di-GMP) is a circular RNA dinucleotide that functions as a second messenger in diverse species of bacteria to trigger wide-ranging physiological changes, including cell differentiation, conversion between motile and biofilm lifestyles, and virulence gene expression. However, the mechanisms by which cyclic di-GMP regulates gene expression have remained a mystery. We found that cyclic di-GMP in many bacterial species is sensed by a riboswitch class in messenger RNA that controls the expression of genes involved in numerous fundamental cellular processes. A variety of cyclic di-GMP regulons are revealed, including some riboswitches associated with virulence gene expression, pilus formation, and flagellum biosynthesis. In addition, sequences matching the consensus for cyclic di-GMP riboswitches are present in the genome of a bacteriophage.

The second messenger cyclic di-GMP (*1*–*4*) (fig. S1) is formed from two guanosine-5′-triphosphate molecules by diguanylate cyclase (DGC) enzymes. Once formed, the compound is degraded selectively by phosphodiesterase (PDE) enzymes that contain either EAL or HD-GYP amino acid domains (*5*). The activities of these synthesis and degradation enzymes are triggered by various stimuli and modulate cellular cyclic di-GMP concentrations (*4*) to signal physiological changes. Cyclic di-GMP can be bound by some DGC proteins to allosterically repress its own synthesis (*6*, *7*). The only other protein targets known are *Gluconacetobacter xylinus* cellulose synthase (*8*, *9*), *Pseudomonas aeruginosa* PelD protein (*10*), and PilZ domain proteins (*11*). However, cyclic di-GMP binding presumably affects only the pathways in which these proteins participate, and therefore these interactions cannot fully explain its global cellular effects (*3*, *12*).

It has been hypothesized (*3*) that the existence of cyclic di-GMP riboswitches could explain how

---

[1] Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA. [2] Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520, USA. [3] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

*To whom correspondence should be addressed. E-mail: ronald.breaker@yale.edu

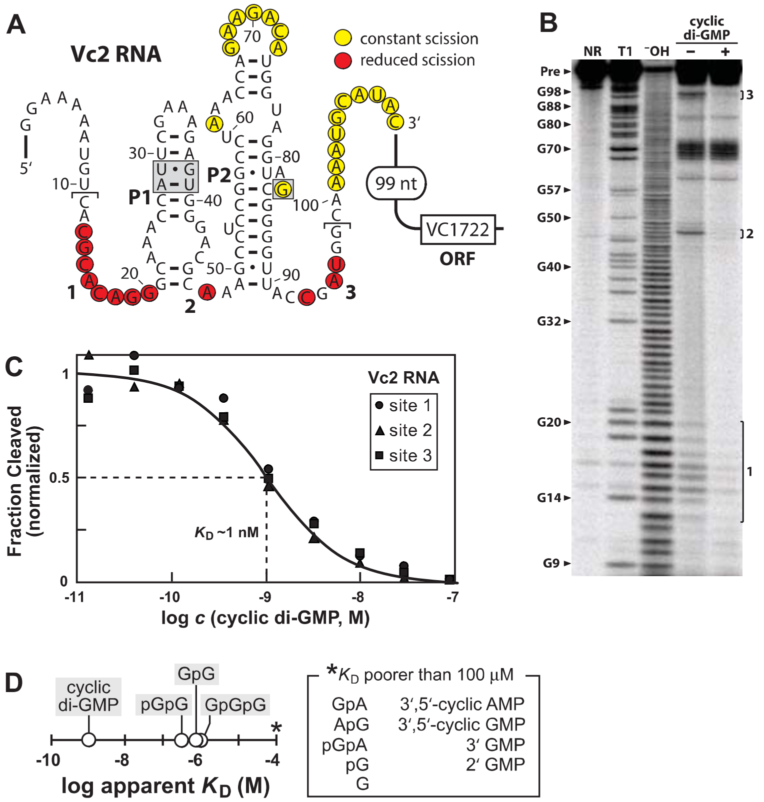


Fig. 1. A cyclic di-GMP aptamer from *V. cholerae*. (**A**) Sequence and structure of the Vc2 RNA from *V. cholerae* chromosome 2, and its proximity to the ORF of *VC1722*. Nucleotides shown correspond to the 110 Vc2 RNA construct. Bold numbers identify regions of ligand-mediated structure modulation as observed in (B). Brackets identify the minimal 5′ or 3′ terminus (when the opposing terminus for 110 Vc2 RNA is retained) that exhibits structural modulation when tested with 10 nM cyclic di-GMP. Nucleotides in shaded boxes were mutated for studies depicted in Fig. 2A. Predicted G and U wobble pairs are identified with dots. (**B**) Polyacrylamide gel electrophoresis of RNA products generated by in-line probing of 5′ ^{32}P-labeled 110 Vc2 RNA. NR, no reaction; T1, partial digest with RNase T1; ⁻OH, partial digest with alkali. RNA was incubated in the absence (–) or presence (+) of 100 µM cyclic di-GMP. (**C**) Plot of the normalized fraction of 110 Vc2 aptamer cleaved versus cyclic di-GMP concentration. Sites of structural modulation are as depicted in (B). (**D**) Comparison of K_D values exhibited by 110 Vc2 aptamer for cyclic di-GMP (fig. S5) and various analogs. G, guanosine; pG, pGpG, and pGpA, 5′ phosphorylated mono- and dinucleotides; GpGpG, trinucleotide; AMP and GMP, adenosine and guanosine monophosphate, respectively.

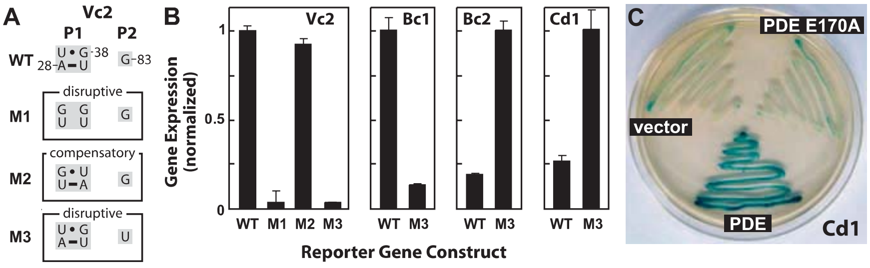


Fig. 2. Representative cyclic di-GMP aptamers are components of gene control elements. (**A**) Reporter fusion constructs carry wild-type (WT) or mutant (M1 through M3) riboswitches from *V. cholerae* (Vc2) in *E. coli*, or carry the equivalent WT and M3 riboswitches from *B. cereus* (Bc1 and Bc2) or *C. difficile* (Cd1) (fig. S6) in *B. subtilis*. (**B**) β-Galactosidase reporter gene assays for reporter fusion constructs described in (A). Maximum Miller units measured for the four aptamer representatives were 436, 47, 5, and 51, respectively. (**C**) *B. subtilis* cells carrying a β-galactosidase reporter construct fused to a wild-type Cd1 riboswitch and transformed with a plasmid lacking or carrying a normal or mutant (E170A) *V. cholerae vieA* gene encoding an EAL PDE. Plate diameter, 85 mm.

this second messenger controls the transcription and translation of many genes. Riboswitches are mRNA domains that control gene expression in response to changing concentrations of their target ligand (*13*, *14*). We have discovered a highly conserved RNA domain called GEMM (*15*) residing upstream of the open reading frames (ORFs) for DGC and PDE proteins in some organisms and residing upstream of some genes that are controlled by cyclic di-GMP. The high conservation and genomic distributions of GEMM RNAs are characteristic of riboswitches.

Most GEMM RNAs conform to one of two similar architectures termed type 1 and type 2 (fig. S2) that are distinguished by the presence of specific tetraloop and tetraloop receptor sequences (*5*). Both types carry two base-paired regions (P1 and P2) that exhibit extensive covariation in the 503 representatives identified (*15*). The genome of the pathogenic bacterium *Vibrio cholerae* carries two sequences for type 1 GEMM RNAs (fig. S3). One (Vc1) resides upstream of the *gbpA* gene, and a second (Vc2) resides upstream of a gene (*VC1722*) homologous to *tfoX* of *V. cholerae* (*15*, *16*).

Biochemical and genetic analyses were carried out with both *V. cholerae* GEMM RNAs to determine whether they function as aptamers for cyclic di-GMP. A 110-nucleotide Vc2 RNA construct (110 Vc2) (Fig. 1A) was subjected to in-line probing (*17*) in the absence or presence of 100 µM cyclic di-GMP (Fig. 1B). The changing patterns of spontaneous cleavage at the base of stems P1 and P2 suggest that the conserved nucleotides near the regions undergoing structural modulation (labeled 1 through 3) are important for ligand binding. Similar results were obtained for a construct encompassing the Vc1 RNA and for representative cyclic di-GMP aptamers from *Bacillus cereus* and *Clostridium difficile*. Despite the constrained structure of the second messenger, cyclic di-GMP exhibited a half-life for spontaneous degradation of no poorer than ~150 days under in-line probing conditions (fig. S4). Therefore, the biologically relevant ligand for this aptamer class most likely is cyclic di-GMP and not one of the breakdown products of this second messenger.

The 110 Vc2 RNA appeared to form a one-to-one saturable complex with a dissociation constant (K_D) of ~1 nM for cyclic di-GMP (Fig. 1C and fig. S5). This interaction was nearly three orders of magnitude tighter than the affinity measured for cyclic di-GMP binding by an *Escherichia coli* PilZ protein domain [K_D = 840 nM (*11*)]. Furthermore, various analogs of the second messenger were strongly discriminated against by the Vc2 aptamer (Fig. 1D). The linear breakdown product of cyclic di-GMP by EAL PDEs (*1*–*3*), pGpG, was bound less tightly by the aptamer by nearly three orders of magnitude. Similarly, the pG (5′ GMP) product of HD-GYP PDEs was not bound by the aptamer even at a concentration five orders of magnitude higher than the K_D for cyclic di-GMP.

Bacterial riboswitches typically carry a conserved aptamer located immediately upstream of a less well-conserved expression platform. Most expression platforms in bacteria form structures that control transcription termination or translation initiation (*18*). Similar architectures are associated with many GEMM RNA representatives (*5*), which suggests that cyclic di-GMP aptamers are likely to be components of riboswitches. Four 5′ untranslated regions from *V. cholerae*, *B. cereus*, and *C. difficile* were examined for riboswitch function in cells. DNAs encompassing Vc2, its predicted native promoter, and the adjoining expression platform were used to prepare translational fusions with the *E. coli lacZ* reporter gene. Transformed *E. coli* cells carrying the aptamer-reporter fusion constructs were grown in liquid medium and assayed for β-galactosidase activity.

The reporter construct carrying the wild-type Vc2 aptamer (Figs. 1A and 2A) exhibited a high level of gene expression (Fig. 2B). In contrast, constructs M1 and M3 that carried mutations disrupting P1 or altering an otherwise strictly conserved nucleotide in the P2 bulge expressed the reporter gene at less than 10% of the wild-type level. Furthermore, an aptamer that carried four compensatory mutations (M2) that restore the structure of P1 exhibited near–wild-type gene expression. These results parallel the ligand-binding activities of these RNAs and suggest that the riboswitch integrating the Vc2 aptamer functions as an "on" switch. Similarly, transcriptional fusions were prepared for cyclic di-GMP riboswitches from *B. cereus* (Bc1 and Bc2) and *C. difficile* (Cd1) (fig. S6) and cloned into *B. subtilis*. The expression patterns for wild-type and equivalent M3 variants were consistent with "on" switch function for Bc1 and "off" switch function for Bc2 and Cd1 (Fig. 2B).

Cyclic di-GMP riboswitch modulation of gene expression was established by examining "off" switch action of the Cd1 RNA. The cellular concentration of the second messenger was manipulated by expression of *V. cholerae* VieA, an EAL PDE that is expected to lower the cellular concentration of cyclic di-GMP by promoting hydrolysis to its linear pGpG form (*19*). The Cd1 riboswitch is located in the 5′ untranslated region of a large operon that encodes the proteins required to construct the entire flagella apparatus of this species (table S1). This arrangement suggests that some bacteria use a riboswitch to sense changes in cyclic di-GMP concentrations and regulate organelle biosynthesis to control motile versus sessile lifestyle transformations.

Transformed *B. subtilis* cells carrying a transcriptional fusion of the wild-type Cd1 riboswitch with *lacZ*, and grown on agar plates with X-gal, exhibited low β-galactosidase activity when cotransformed with either the empty plasmid vector or the vector coding for the inactive $\text{Glu}^{170} \rightarrow$ Ala (E170A) VieA mutant (Fig. 2C). In contrast, robust β-galactosidase activity was evident in cells cotransformed with the reporter construct and the vector coding for VieA PDE. A similar trend was observed when these transformed cells were assayed from liquid culture, whereas a reporter construct carrying the M3 mutation in the Cd1 aptamer remained largely unaffected by VieA activity (fig. S7). Moreover, cyclic di-GMP caused transcription termination of the Cd1 RNA in an in vitro transcription assay (fig. S8). These observations suggest that *C. difficile* uses the Cd1 riboswitch to regulate transcription of the flagellar protein operon to control motility in response to cyclic di-GMP signaling.

The *VC1722* gene associated with the Vc2 riboswitch codes for a protein similar to a known transcription factor that is important for competence (*16*). A *V. cholerae* mutant producing a rugose phenotype was found to have elevated cyclic di-GMP levels as well as higher expression of the *VC1722* mRNA (*20*, *21*). This is consistent with our data indicating that the Vc2 riboswitch associated with the *VC1722* mRNA is a genetic "on" switch that yields higher gene expression when cyclic di-GMP concentrations are elevated (Fig. 2C).

The remaining *V. cholerae* cyclic di-GMP riboswitch (Vc1; fig. S3) is associated with the *gbpA* gene, which codes for a sugar-binding protein reported to be the key determinant permitting the bacterium to colonize mammalian intestines, leading to cholera disease (*22*). It has been shown that *V. cholerae* lowers its cyclic di-GMP levels when colonizing mammalian host intestines (*19*). The *V. cholerae vieA* gene used in our study to demonstrate Cd1 riboswitch response to changing cyclic di-GMP levels (Fig. 2C) is known to be essential for bacterial infection. Thus, it is possible that a reduction in cyclic di-GMP levels brought about by the action of VieA is sensed by the Vc1 riboswitch to facilitate expression of GbpA and *V. cholerae* infection.

Some organisms have a strikingly high number of cyclic di-GMP riboswitch representatives (fig. S3) (*5*). *Geobacter uraniumreducens*, with 30 representatives identified, has the largest number of cyclic di-GMP aptamer RNAs among bacterial species whose genomes have been sequenced. These are distributed upstream of 25 different transcriptional units, with five RNAs carrying tandem cyclic di-GMP aptamers. Also intriguing is the identification of cyclic di-GMP riboswitch representatives residing within PhiCD119 bacteriophage DNA that is integrated within the *C. difficile* genome. The riboswitch sequence, located within the lysis module of the bacteriophage genome, is also evident in DNA packaged into bacteriophage particles (*23*). Although more than 20 metabolite-sensing riboswitch classes have been reported, and thousands of representatives of these classes have been identified, the cyclic di-GMP–binding RNAs are the only examples of bacteriophage-associated riboswitches found after exhaustive bioinformatics searches. Perhaps viruses have little need for sensing fundamental metabolic products, but might gain an evolutionary advantage by monitoring the physiological transformations of bacterial cells brought about by changing concentrations of the second messenger cyclic di-GMP.

References and Notes

1. U. Jenal, J. Malone, *Annu. Rev. Genet.* **40**, 385 (2006).
2. R. P. Ryan, Y. Fouhy, J. F. Lucey, J. Maxwell Dow, *J. Bacteriol.* **188**, 8327 (2006).
3. R. Tamayo, J. T. Pratt, A. Camilli, *Annu. Rev. Microbiol.* **61**, 131 (2007).
4. R. Simm, M. Morr, A. Kader, M. Nimtz, U. Römling, *Mol. Microbiol.* **53**, 1123 (2004).
5. See supporting material on *Science* Online.
6. C. Chan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17084 (2004).
7. P. Wassmann *et al.*, *Structure* **15**, 915 (2007).
8. P. Ross *et al.*, *FEBS Lett.* **186**, 191 (1985).
9. P. Ross *et al.*, *Nature* **325**, 279 (1987).
10. V. T. Lee *et al.*, *Mol. Microbiol.* **65**, 1474 (2007).
11. D. A. Ryjenkov, R. Simm, U. Römling, M. Gomelsky, *J. Biol. Chem.* **281**, 30310 (2006).
12. A. J. Wolfe, K. L. Visick, *J. Bacteriol.* **190**, 463 (2008).
13. W. C. Winkler, R. R. Breaker, *Annu. Rev. Microbiol.* **59**, 487 (2005).
14. R. R. Breaker, *Science* **319**, 1795 (2008).
15. Z. Weinberg *et al.*, *Nucleic Acids Res.* **35**, 4809 (2007).
16. K. L. Meibom, M. Blokesch, N. A. Dolganov, C.-Y. Wu, G. K. Schoolnik, *Science* **310**, 1824 (2005).
17. G. A. Soukup, R. R. Breaker, *RNA* **5**, 1308 (1999).
18. J. E. Barrick, R. R. Breaker, *Genome Biol.* **8**, R239 (2007).
19. R. Tamayo, A. D. Tischler, A. Camilli, *J. Biol. Chem.* **280**, 33324 (2005).
20. B. Lim, S. Beyhan, J. Meir, F. H. Yildiz, *Mol. Microbiol.* **60**, 331 (2006).
21. S. Beyhan, F. H. Yildiz, *Mol. Microbiol.* **63**, 995 (2007).
22. T. J. Kirn, B. A. Jude, R. K. Taylor, *Nature* **438**, 863 (2005).
23. R. Govind, J. A. Fralick, R. D. Rolfe, *J. Bacteriol.* **188**, 2568 (2006).

24. We thank D. Rivera for assistance with experiments analyzing cyclic di-GMP stability and integrity. We also acknowledge the efforts of the FT-ICR Mass Spectrometry Resource of the Keck Biotechnology Resource Laboratory, and we thank N. Carriero and R. Bjornson for assisting with our use of the Yale Life Sciences High Performance Computing Center (NIH grant RR19895-02). Supported by NIH grants R33 DK07027 and GM 068819 (R.R.B.), National Institute of General Medical Sciences grant T32GM007223 (E.R.L.), an NSF predoctoral fellowship (J.N.K.), and National Heart, Lung, and Blood Institute grant HV28186. RNA research in the Breaker laboratory is also supported by the Howard Hughes Medical Institute. R.R.B. is a cofounder and member of the scientific advisory board of BioRelix. N.S., Z.W., and R.R.B. are listed as inventors on several patent applications related to riboswitch science.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/411/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 and S2
References

23 April 2008; accepted 12 June 2008
10.1126/science.1159519

Bottom-Up Dependent Gating of Frontal Signals in Early Visual Cortex

Leeland B. Ekstrom,[1,2,3] Pieter R. Roelfsema,[4,5] John T. Arsenault,[1,6] Giorgio Bonmassar,[1] Wim Vanduffel[1,6,7]*

The frontal eye field (FEF) is one of several cortical regions thought to modulate sensory inputs. Moreover, several hypotheses suggest that the FEF can only modulate early visual areas in the presence of a visual stimulus. To test for bottom-up gating of frontal signals, we microstimulated subregions in the FEF of two monkeys and measured the effects throughout the brain with functional magnetic resonance imaging. The activity of higher-order visual areas was strongly modulated by FEF stimulation, independent of visual stimulation. In contrast, FEF stimulation induced a topographically specific pattern of enhancement and suppression in early visual areas, but only in the presence of a visual stimulus. Modulation strength depended on stimulus contrast and on the presence of distractors. We conclude that bottom-up activation is needed to enable top-down modulation of early visual cortex and that stimulus saliency determines the strength of this modulation.

Contemporary hypotheses propose that feedback signals from areas in frontal and parietal cortex exert control over the processing of incoming visual information (*1*–*5*). Several models suggest that these signals are gated by bottom-up stimulation (*6*–*9*). In these models, feedback signals only influence neurons activated by visual input, just as has been observed for attentional effects, which are known to be strongest for neurons well driven by a visual stimulus (*10*–*12*). No causal evidence exists, however, to support these hypotheses, with the exception of area V4, where feedback effects evoked by stimulation of the FEF are most pronounced for neurons strongly activated by a visual stimulus (*13*). To (i) test these models of bottom-up dependent gating of frontal signals on a whole-brain scale, (ii) investigate the impact of increased FEF activity on visually driven responses throughout occipito-temporal cortex, (iii) examine the spatial organization of any observed modulations, and (iv) investigate the effects of visual saliency on such modulations, we developed a combination of functional magnetic resonance imaging (fMRI) and chronic electrical microstimulation (EM) in awake, behaving monkeys.

In our first experiment, the goal was to detect the functional consequences of EM of the FEF in the absence of a visual stimulus, using stimulation levels below those needed to evoke saccades. We first obtained anatomical (fig. S1B) and behavioral evidence (Fig. 1A and fig. S2) in two monkeys that several chronically implanted microelectrodes were positioned in the FEF. Before each fMRI experiment, we stimulated these electrodes inside the MR scanner to determine the threshold needed to evoke saccades and to identify the saccade end point, or movement field (MF), of each FEF stimulation site (Fig. 1A and fig. S2). During the actual fMRI experiment, the monkeys carried out a passive fixation task while we alternated between epochs of no-EM and epochs of EM, at a stimulation level of 50% of the saccade-inducing amplitude. The use of this method in awake animals allowed us to titrate the stimulation to functionally relevant levels (*14*), an advantage compared to a previous study in anesthetized animals (*15*).

The left column of Fig. 1B shows *t*-score maps of regions with increased fMRI activity caused by FEF-EM overlaid on coronal, T1-weighted sections. Focal increases in fMRI activity were observed at the site of stimulation and across the brain, in regions known to be connected to the FEF (*16*–*20*) [Figs. S3 and S4; see (*14*) for a full list of areas]. To illustrate the amplitude of EM-induced fMRI effects in five representative regions, we plotted percentage change in MR signal relative to the no-EM condition (Fig. 1C). In comparison with classical tracer studies (*17*, *18*), we obtained virtually identical results with EM-fMRI (figs. S3 and S4). This close correspondence demonstrates the precision of the technique and suggests that we primarily activated regions monosynaptically connected to the stimulation site. At the statistical threshold used ($P < 0.05$, corrected), we observed no negative EM-induced activations in visual cortex. The right column of Fig. 1B represents a replication of the experiment after 1 month, show-

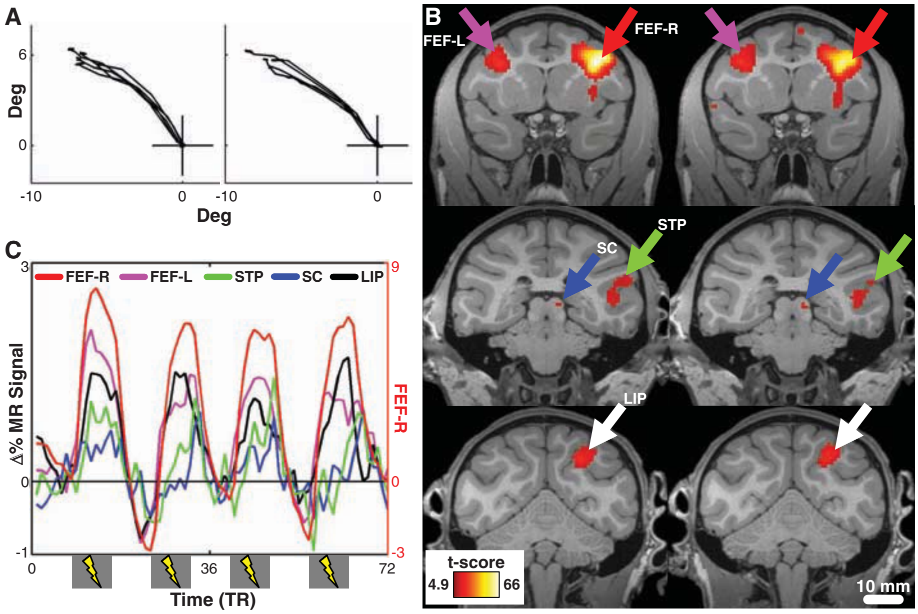


Fig. 1. fMRI activity induced by right FEF-EM. (**A**) Eye traces separated by 1 year showing saccades evoked by suprathreshold stimulation of one electrode. (**B**) Coronal slices showing activity (*t*-score maps) from the contrast EM versus no-EM (MM1, $P < 0.05$, corrected; R/L, right/left; see table S1 for anatomical abbreviations). Columns represent test-retest separated by 1 month. (**C**) Time courses of percentage change in MR signal in areas in (B). Note the secondary *y* axis for the FEF-R data. Gray bars (*x* axis) indicate stimulation epochs.

---

[1] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA. [2] Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [3] Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [4] Department of Vision and Cognition, Netherlands Institute for Neuroscience, an institute of the Royal Netherlands Academy of Arts and Sciences, Meibergdreef 47, 1105 BA, Amsterdam, Netherlands. [5] Department of Experimental Neurophysiology, Center for Neurogenomics and Cognitive Research, Vrije Universiteit, de Boelelaan 1085, 1081 HV, Amsterdam, Netherlands. [6] Laboratorium voor Neuro- en Psychofysiologie, K. U. Leuven Medical School, Campus Gasthuisberg, 3000 Leuven, Belgium. [7] Department of Radiology, Harvard Medical School, Charlestown, MA 02129, USA.

*To whom correspondence should be addressed. E-mail: wim@nmr.mgh.harvard.edu

ing the reproducibility of the results even for small foci, such as those seen in the superior colliculus (*14*).

Our first experiment indicated that FEF-EM increased fMRI activity in higher-order visual areas known to be directly connected to the FEF. If feedback effects are gated by visual stimulation, however, one also predicts FEF-EM effects in visual areas separated from the FEF by multiple synapses, in the presence of a visual stimulus. In a second experiment, we therefore placed high-contrast, colored, moving gratings in the MFs of the FEF stimulation sites under passive viewing conditions (fig. S5A) (*14*) and measured the fMRI response to EM-only, visual-only (V), and combined visual-EM (VEM) stimulation, relative to a fixation-only (F) condition (i.e., a 2 by 2 factorial design with factors EM and visual stimulation). To investigate the net influence of visual stimulation on FEF-EM effects, we compared EM minus fixation (Fig. 2A) to VEM minus V (Fig. 2C) in all visually driven voxels for a number of cortical areas; visually driven activity without EM is shown in Fig. 2B. Visually driven voxels in areas directly connected to the FEF (*16*–*20*), such as the lateral intraparietal area (LIP) and several areas within the superior temporal sulcus, showed an EM-driven increase in fMRI activity that was relatively independent of the presence of a visual stimulus (compare Figs. 2A and 2C). In contrast, visual stimulation unveiled a significant influence of FEF-EM on the activity of early visual areas (V1, V2, V3, and V4) (*14*). Figure 2D isolates the effect of the visual stimulus on FEF-EM by subtracting the activity evoked by EM in the absence of a visual stimulus (Fig. 2A) from the activity evoked in the presence of a visual stimulus (Fig. 2C). Visual stimulation enabled the effects of FEF microstimulation to reach early visual areas, including V1, which is not monosynaptically connected to the FEF.

We next investigated the spatial patterns of increased or suppressed fMRI activity in each of the areas, because the absence of an overall net effect might nevertheless be associated with a pattern of balanced inhibitory and excitatory effects. Figure 3C shows the cortical regions where we observed positive and negative interactions between visual stimulation and FEF-EM in our 2 by 2 factorial design (Fig. 3A; see Fig. 3B for the stimulus positions used). Voxels shown in yellow-orange are visually driven; visually driven voxels with a positive or negative interaction are shaded green or blue, respectively (fig. S6; see fig. S7 for MM2). The spatial pattern of these interactions was heterogeneous in most visual areas: FEF-EM mostly amplified visual activity in subsets of voxels (green) adjacent to those maximally driven by the visual stimuli alone (yellow-orange), whereas it tended to have little influence or even suppressed the response (blue) in voxels strongly driven by the visual stimuli.

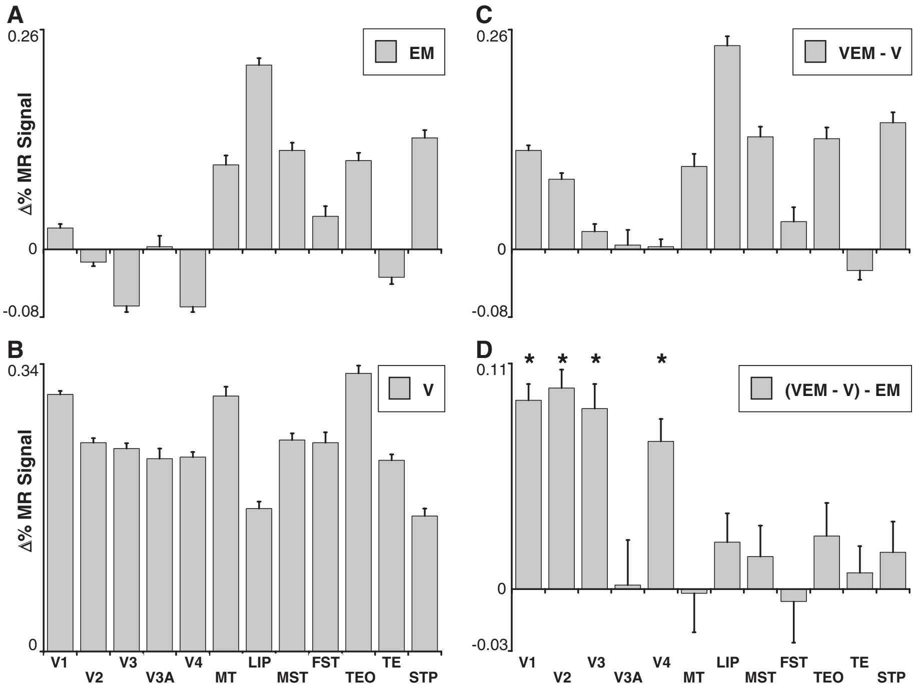


Fig. 2. Gating of FEF-EM effects by a visual stimulus. Mean percentage change in MR signal relative to fixation-only (MM1 and MM2) for all visually driven voxels ($P < 0.05$, uncorrected) in 12 visual areas for (**A**) EM and (**B**) V epochs, (**C**) the difference between VEM and V epochs, and (**D**) the interaction (VEM-V-EM). Error bars denote SEM; * indicates a significant difference between the VEM-V and EM distributions ($P < 10^{-17}$, two-sample, two-tailed *t* test), revealing that on average voxels in early visual areas show a larger EM response in the presence of visual stimuli than in the absence. Higher-order visual areas show an EM response both with and without visual stimulation.

We further quantified these interactions between the visual stimuli and FEF-EM at the level of individual visually driven voxels by comparing the percentage change in MR signal in V epochs with the difference between VEM and EM epochs (Fig. 3D and fig. S8). The visual voxels boosted by FEF-EM were weakly driven by the visual stimuli alone, whereas the voxels that were strongly driven by the visual stimuli were unaffected (both subjects) or even suppressed (mostly MM1) by FEF-EM. We carried out several analyses to exclude the possibility that the modulation of visually driven activity by FEF-EM was due to differences in eye position or saccade rate (*14*) (table S2 and figs. S9 and S10).

Modulatory effects from the FEF in visual cortex could be spatially nonspecific or could require the precise alignment of a visual stimulus with the FEF MF (*13*). Our third experiment compared the effects of incongruent epochs (VEM-I), in which EM was applied to a particular FEF MF while a visual stimulus was presented in another, nonstimulated MF (separated by 6.5 to 13.7 degrees), to congruent epochs that were identical to the VEM condition of experiment two (Fig. 4A and fig. S5B). Net EM and visual stimulation were exactly matched between congruent and incongruent epochs. In Fig. 4B, we show a portion of flattened right occipital cortex overlaid with a t-score map of visually driven fMRI activity. The spatial pattern supports that seen in experiment two—in general, voxels adjacent to those maximally driven by the visual stimulus alone showed more fMRI activity during congruent than during incongruent FEF-EM (green). An analysis at the level of individual voxels confirmed that most voxels were better activated during VEM than during VEM-I, especially in motion-sensitive areas such as MT, MST, and FST (Fig. 4C and fig. S11) (*14*, *21*).

To assess more directly whether FEF-EM mainly influences the activity of nonoptimally driven voxels, our next two experiments manipulated the saliency or strength of the visual response. In experiment four, we placed one stimulus in the FEF-MF and added three distractors (D) in the opposite hemifield (fig. S12). For comparative purposes (*13*), we focus here on visually driven regions in V4 only. As expected, the distractors did not evoke a response in V4 representing the opposite visual field (Fig. 5A). Microstimulation of the FEF in the presence of distractors, however, produced the largest response for the stimulus in the FEF MF, in agreement with a previous study (*13*). The interaction between distractors and FEF-EM was significant ($P < 1.4 \times 10^{-5}$). In a fifth experiment, we varied the luminance contrast of a single stimulus placed in the FEF MFs and observed significant enhancement effects of FEF-EM for low-contrast stimuli only. These results are in accordance with experiment two, where we obtained the strongest modulation in voxels with a weaker visual response.

We have demonstrated spatially specific, causal interactions between activity in area FEF and many areas of the visual cortex. Signals from frontal cortex activated higher-order areas directly connected to the FEF irrespective of visual stimulation, mimicking attention-driven baseline shifts in activity (*14*, *22*). In the presence of visual stimulation, modulations were observed at even the earliest levels of visual cortex, including area V1, which may account for previous findings that the effects of FEF-EM are retinotopically highly specific (*23*). These effects are likely trans-synaptic, because these early areas do not receive direct connections from the FEF and were only observed when the neurons were congruently activated by a visual stimulus. One interpretation is that visual stimulation opens feedback pathways closed in the absence of stimuli (*6*–*9*), allowing these frontal signals to propagate from higher to earlier visual cortical areas. The effect of FEF-EM thus resembles spatial attention, which interacts

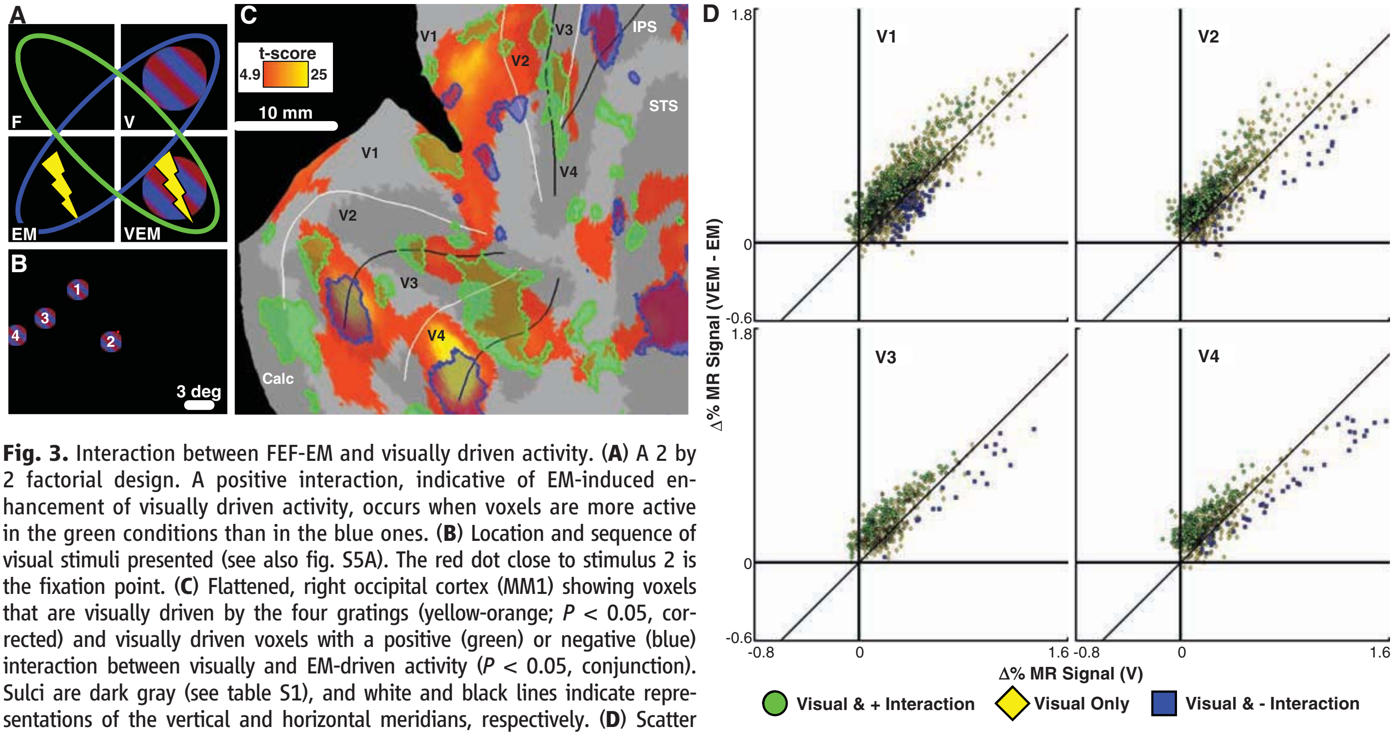


Fig. 3. Interaction between FEF-EM and visually driven activity. (**A**) A 2 by 2 factorial design. A positive interaction, indicative of EM-induced enhancement of visually driven activity, occurs when voxels are more active in the green conditions than in the blue ones. (**B**) Location and sequence of visual stimuli presented (see also fig. S5A). The red dot close to stimulus 2 is the fixation point. (**C**) Flattened, right occipital cortex (MM1) showing voxels that are visually driven by the four gratings (yellow-orange; $P < 0.05$, corrected) and visually driven voxels with a positive (green) or negative (blue) interaction between visually and EM-driven activity ($P < 0.05$, conjunction). Sulci are dark gray (see table S1), and white and black lines indicate representations of the vertical and horizontal meridians, respectively. (**D**) Scatter plots (MM1 and MM2) of voxels in areas V1, V2, V3, and V4 (see also fig. S8) showing percentage change in MR signal for V epochs relative to fixation-only epochs (x axis) and the difference between VEM and EM epochs (y axis). Color code matches (C) (yellow now at $P < 0.05$, uncorrected). Points close to the y axis are weakly visually driven; voxels enhanced by FEF-EM (green) are mainly clustered near the y axis, whereas strongly visually driven voxels are either unaffected or suppressed (blue).

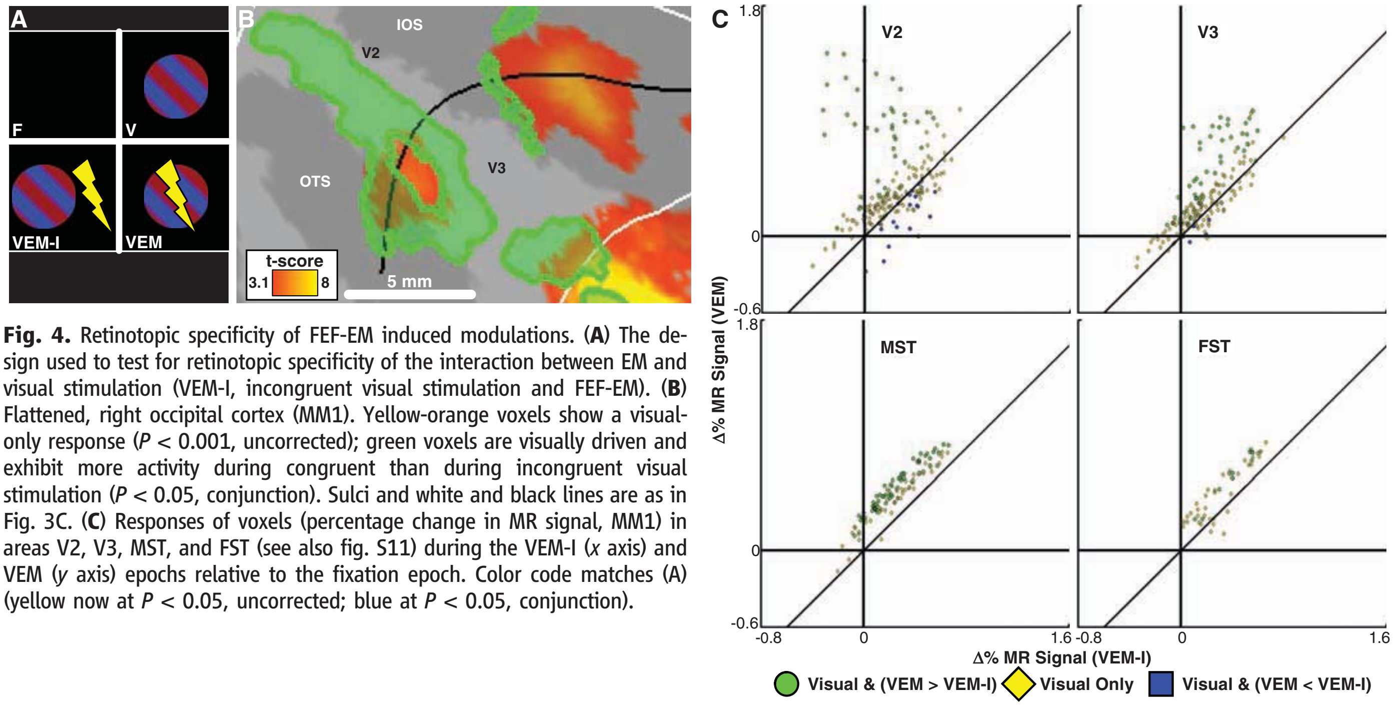


Fig. 4. Retinotopic specificity of FEF-EM induced modulations. (**A**) The design used to test for retinotopic specificity of the interaction between EM and visual stimulation (VEM-I, incongruent visual stimulation and FEF-EM). (**B**) Flattened, right occipital cortex (MM1). Yellow-orange voxels show a visual-only response ($P < 0.001$, uncorrected); green voxels are visually driven and exhibit more activity during congruent than during incongruent visual stimulation ($P < 0.05$, conjunction). Sulci and white and black lines are as in Fig. 3C. (**C**) Responses of voxels (percentage change in MR signal, MM1) in areas V2, V3, MST, and FST (see also fig. S11) during the VEM-I (x axis) and VEM (y axis) epochs relative to the fixation epoch. Color code matches (A) (yellow now at $P < 0.05$, uncorrected; blue at $P < 0.05$, conjunction).

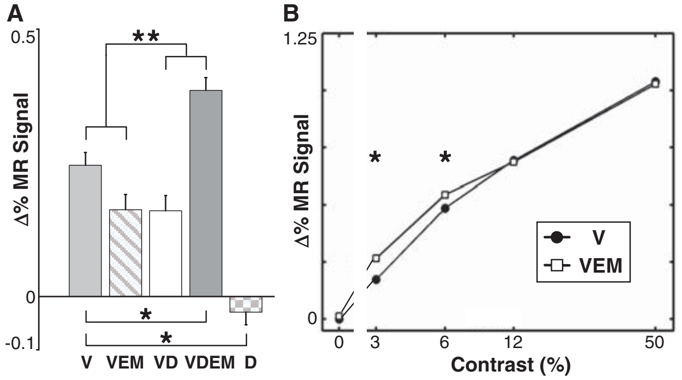


Fig. 5. Effects of distractor stimuli and stimulus contrast. (**A**) Mean percentage change in MR signal relative to fixation-only in visually driven voxels in V4 (MM1; $P <$ 0.05, uncorrected) during V, VEM, visual + distractor (VD), visual + distractor + EM (VDEM), and distractor-only (D) epochs. Two FEF MFs were stimulated (fig. S12). Error bars denote SEM. * denotes a significant difference between the pair of conditions indicated ($P < 0.05$, two-sample, two-tailed t test); ** indicates that the interaction between distractors and EM is significant ($P < 1.4 \times 10^{-5}$). (**B**) Mean percentage change in MR signal relative to the fixation epoch in visually driven voxels in V4 (MM1 and MM2) as a function of stimulus luminance contrast. Only one visual stimulus was used per session (fig. S5A), with the data combined across all four MF locations from each subject. The highest contrast stimulus was used as a localizer ($P < 0.05$, uncorrected) (*14*). Error bars, smaller than the plot symbols, denote SEM. * denotes a significant difference between the VEM and V conditions at a given contrast ($P < 0.05$, two-sample, two-tailed t test).

with visual stimuli in a comparable, multiplicative manner (*10–12*).

An unexpected finding was that the voxels most strongly enhanced by FEF-EM were adjacent to the voxels with the strongest visual response, which were themselves unaffected or even suppressed by FEF-EM. These results support a recent model for the effects of feedback connections proposing that feedback and horizontal connections mediate a contrast-dependent inhibition of a central zone in the next lower area while exciting the near surround (*24*). Strong effects of FEF-EM in weakly driven voxels also dovetails with previous findings that attentional effects and the effects of FEF microstimulation on neuronal activity in area V4 are most pronounced in the presence of competitive distractors that reduce the strength of neuronal responses (*13*, *25*). As another means to reduce the activity of V4 neurons, we lowered the luminance contrast of the visual stimulus. We observed that FEF-EM evoked stronger effects in low-contrast rather than high-contrast stimuli. This result is reminiscent of a nonproportional scaling of the contrast response function in area V4, with stronger positive modulations for low-contrast stimuli, as seen in previous spatial attention studies [(*26*), but see (*27*)]. All these results taken together make it tempting to speculate that FEF-EM engages similar neuronal circuits as spatial attention. Speculation aside, the present results clearly strengthen past observations that structures involved in generating eye movements (*1*, *13*, *28–30*) are well suited to modulate sensory-driven activity in a topographically specific manner.

References and Notes

1. G. Rizzolatti, L. Riggio, I. Dascola, C. Umilta, *Neuropsychologia* **25**, 31 (1987).
2. S. Kastner, L. G. Ungerleider, *Annu. Rev. Neurosci.* **23**, 315 (2000).
3. M. Corbetta, G. L. Shulman, *Nat. Rev. Neurosci.* **3**, 201 (2002).
4. T. Moore, K. M. Armstrong, M. Fallah, *Neuron* **40**, 671 (2003).
5. F. H. Hamker, *Cereb. Cortex* **15**, 431 (2005).
6. K. Fukushima, *IEEE Computer* **21**, 65 (1988).
7. S. Grossberg, *Spat. Vis.* **12**, 163 (1999).
8. F. van der Velde, M. de Kamps, *J. Cogn. Neurosci.* **13**, 479 (2001).
9. P. R. Roelfsema, *Annu. Rev. Neurosci.* **29**, 203 (2006).
10. S. Treue, J. C. Martinez Trujillo, *Nature* **399**, 575 (1999).
11. C. J. McAdams, J. H. Maunsell, *J. Neurosci.* **19**, 431 (1999).
12. J. C. Martinez-Trujillo, S. Treue, *Curr. Biol.* **14**, 744 (2004).
13. T. Moore, K. M. Armstrong, *Nature* **421**, 370 (2003).
14. Materials and methods are available as supporting material on *Science* Online.
15. A. S. Tolias *et al.*, *Neuron* **48**, 901 (2005).
16. M. F. Huerta, L. A. Krubitzer, J. H. Kaas, *J. Comp. Neurol.* **253**, 415 (1986).
17. M. F. Huerta, L. A. Krubitzer, J. H. Kaas, *J. Comp. Neurol.* **265**, 332 (1987).
18. J. D. Schall, A. Morel, D. J. King, J. Bullier, *J. Neurosci.* **15**, 4464 (1995).
19. G. B. Stanton, M. E. Goldberg, C. J. Bruce, *J. Comp. Neurol.* **271**, 473 (1988).
20. G. B. Stanton, C. J. Bruce, M. E. Goldberg, *J. Comp. Neurol.* **353**, 291 (1995).
21. W. Vanduffel *et al.*, *Neuron* **32**, 565 (2001).
22. C. L. Colby, J. R. Duhamel, M. E. Goldberg, *J. Neurophysiol.* **76**, 2841 (1996).
23. K. M. Armstrong, J. K. Fitzgerald, T. Moore, *Neuron* **50**, 791 (2006).
24. L. Schwabe, K. Obermayer, A. Angelucci, P. C. Bressloff, *J. Neurosci.* **26**, 9117 (2006).
25. J. H. Reynolds, L. Chelazzi, R. Desimone, *J. Neurosci.* **19**, 1736 (1999).
26. J. H. Reynolds, T. Pasternak, R. Desimone, *Neuron* **26**, 703 (2000).
27. T. Williford, J. H. Maunsell, *J. Neurophysiol.* **96**, 40 (2006).
28. M. Corbetta *et al.*, *Neuron* **21**, 761 (1998).
29. J. Cavanaugh, R. H. Wurtz, *J. Neurosci.* **24**, 11236 (2004).
30. C. C. Ruff *et al.*, *Curr. Biol.* **16**, 1479 (2006).
31. We thank H. Deng for animal training and care; M. Khachaturian, H. Kolster, J. Mandeville, G. Madan, T. van Kerkoerle, and L. Wald for technical assistance; R. Tootell and B. Rosen for advice and support; and R. Buckner, G. Orban, and J. Sharma for valuable comments. This work received support from a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship, Human Frontier Science Program Organization, Geneeskundige Stichting Koningin Elisabeth, Inter University Attraction Pole 5/04, Excellence Financing/05/014, Geconcerteerde Onderzoeks Actie 2005/18, Fonds Wetenschappelijk Onderzoek–Vlaanderen G.0.622.08, European Union grant FP7/2007-2013 # F2-2008-200728, R01-EB000790. The Martinos Center is supported by National Center for Research Resources grant P41RR14075 and the Mind Research Network Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/414/DC1
Materials and Methods
SOM Text
Figs. S1 to S12
Tables S1 and S2
References

26 November 2007; accepted 17 June 2008
10.1126/science.1153276

Evolutionary Origins for Social Vocalization in a Vertebrate Hindbrain–Spinal Compartment

Andrew H. Bass,[1,4,5]* Edwin H. Gilland,[2,4] Robert Baker[3,4]

The macroevolutionary events leading to neural innovations for social communication, such as vocalization, are essentially unexplored. Many fish vocalize during female courtship and territorial defense, as do amphibians, birds, and mammals. Here, we map the neural circuitry for vocalization in larval fish and show that the vocal network develops in a segment-like region across the most caudal hindbrain and rostral spinal cord. Taxonomic analysis demonstrates a highly conserved pattern between fish and all major lineages of vocal tetrapods. We propose that the vocal basis for acoustic communication among vertebrates evolved from an ancestrally shared developmental compartment already present in the early fishes.

Although the genetic basis for human speech receives much attention [e.g., (*1*, *2*)], the fundamental issue of the ancestral origins of neural networks for vocal signaling is essentially unexplored. Social, context-dependent acoustic communication occurs in most of the major vertebrate lineages, including fishes (Fig. 1A). Teleost fish, the most species-rich of all vertebrate groups (*3*), have a simple repertoire of vocalizations complemented by vocal and auditory pathways that are organized similarly to those of amphibians, reptiles, birds, and mammals (*4*). Batrachoidid fish (midshipman and toadfish), in particular, have an expansive vocal-acoustic network, including a rhythmically firing, pacemaker–motor neuron circuit that directly determines the contraction rate of vocal muscles attached to the swim bladder and, in turn, the temporal properties of calls (*5*, *6*) (Fig. 1B and fig. S1; movies S1 to S3). Because batrachoidids also have readily studied larval stages (*7*), they were chosen to investigate the hypothesis that fish and terrestrial vertebrates share an ancestral origin of their vocal motor networks. Here, we show that the vocal systems of fishes and tetrapods develop very similarly in a segment-like region that forms a transitional compartment between the caudal hindbrain and rostral spinal cord.

Comparative evidence shows that the developing hindbrain of vertebrates has eight segments or rhombomeres (*8*). A segmental organization has been shown, in part, by identifying hindbrain reticular neurons that project to the spinal cord (*9*). We identified a similar pattern of reticulospinal pathways in larval batrachoidid fish, labeling the rostral spinal cord with either fluorescent dextran-amines or Alexa biocytin and then visualizing retrogradely filled hindbrain neurons with laser-scanning confocal microscopy (*10*). Larvae had eight rhombomeres (rh1 to rh8), with rh1 to rh7 distinguished by clusters, and rh8 by an elongated column, of reticulospinal neurons (Fig. 2A). Rhombomere 8 is of special interest for two reasons. First, it is proposed to give rise to pattern-generating circuitry in all vocal vertebrates (*11*). Second, rh8 forms a transitional zone with the spinal cord that is typically two to three times the size of each rhombomere in segments 2 to 6 (*8*, *12*), and rh8 is subdivided in teleosts (*9*) and birds (*13*) into multiple, segment-like regions. A segmental pattern for rh8 was revealed here in larval batrachoidids by using fluorescent tracers with nonoverlapping wavelengths to simultaneously map the reticulospinal scaffold and vocal motor neurons [labeled via occipital nerves innervating developing vocal muscle, see (*7*)]. A dense midline column of vocal motor neurons (Fig. 2, B and C) was revealed, lying immediately caudal to a rostral subdivision of rh8's reticulospinal column. The easily delimited vocal motor nucleus (VMN) prefigures its expansive size in postlarval stages (Fig. 2D, also see Fig. 1B). The early larval VMN was more than twice as long as individual rhombomeres within rh2 to rh6, comparable to the two-rhombomere (bimeric) extent of some cranial motor nuclei, namely the trigeminal and abducens nuclei, which span, respectively, rh2 to rh3 and rh5 to rh6 in teleost fish and birds (*8*, *12*, *13*).

To further test the possibility that the VMN in larval fish develops within rh8, its position relative to other neuronal groups was examined. The labeling of vocal and nonvocal muscles with different detxran-amines showed that the caudal half of the VMN was aligned with the rostral portion of a motor column, extending into rostral spinal segments, that innervates pectoral girdle and trunk muscles (PEC and EP, respectively) (Fig. 3, A and B); direct labeling of the occipital nerve roots that innervate these muscles (*7*, *14*) confirmed this pattern (Fig. 3C). We then compared the locations of vocal neurons with those of vagal motor neurons, the caudal extent of which provide a landmark for the hindbrain-spinal transition (*13*). Composite labeling of the vagal-innervated gastrointestinal wall (fig. S1B) and occipital-innervated pectoral girdle muscle, along with vocal neurons, showed two cell groups (Fig. 3, D to F). The gastrointestinal neurons (XMNs) corresponded to a caudal subgroup of vagal motor neurons (*15*) that extended along the rostral half of the VMN, whereas the pectoral neurons (LPs) represented the far rostral pole of the occipital-spinal column (Fig. 3C). An antibody to choline acetyltransferase demarcated the entire vagal motor column in a pattern consistent with the dextran-amine labels (fig. S2), thus reinforcing its identification. The simultaneous mapping of nonvocal and vocal motor neurons led to the hypothesis that the VMN in fish develops in a distinct compartment formed by hindbrain rh8 and rostral spinal cord.

The fully differentiated vocal circuit includes both pacemaker and motor neurons that are readily labeled via transneuronal transport of neurobiotin applied to the vocal branch of the occipital nerve (*6*). Only vocal motor neurons are labeled at the earliest larval stages, followed by pacemaker and prepacemaker neurons at successively later stages (*16*). The developing vocal circuit of larval fish was mapped here relative to the segmental organization of the hindbrain at the earliest stage that includes prepacemaker and pacemaker neurons (*10*). Alignment of a horizontal series of neurobiotin-filled vocal (Fig. 4, A and B) and reticulospinal (Fig. 4, C and D) neurons in larvae

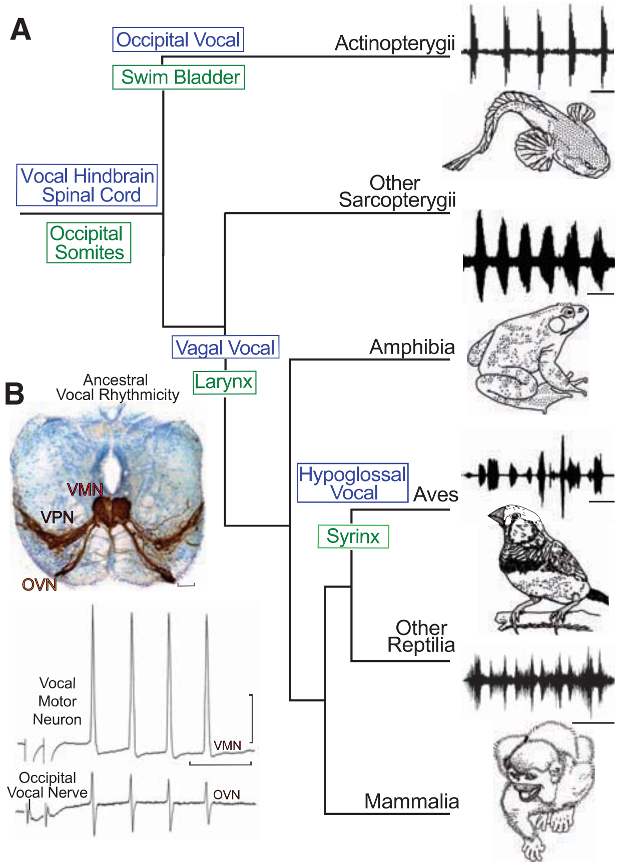


Fig. 1. Evolution of vocal behaviors. (**A**) Cladogram of living bony vertebrates (*3*), with oscillogram of a vocalization from a representative species, shows nodal (ancestral) states for vocal characters (*10*). Vocalizations (top to bottom): midshipman fish agonistic "grunts," bullfrog advertisement call, estrildid finch song, and squirrel monkey cackle. Scale bars (top to bottom) are 500 ms, 1.0 s, 250 ms, and 200 ms. Vocal mechanisms are unknown for lobe-finned fish (other Sarcopterygii), although well-known for nonavian (other) Reptilia (*30*). (**B**) Vocal pacemaker circuit. Among batrachoidid fish (midshipman and toadfish), there is a direct translation between the temporal properties of the vocal circuitry and natural calls (*4*, *5*). (Top) Transverse section at the caudal hindbrain-spinal cord transition of a toadfish shows transneuronal, neurobiotin-labeling of midline vocal motor neurons (VMNs), adjacent pacemaker neurons (VPN), and motor axons exiting via nerve root that gives rise to the occipital vocal nerve (OVN); vocal neurons have extensive lateral processes (*5*, *6*). Scale bar is 100 μm. (Bottom) The rhythmic, oscillatory-like activity of a vocal motor neuron [(top trace) average of four DC-coupled intracellular records] is aligned with occipital nerve activity [(bottom trace) average of four intracranial records] to indicate relative timing [see (*5*) for same temporal pattern of pacemaker neurons]. Response was evoked by midbrain electrical microstimulation in midshipman fish [paired stimulus artifact far left; modified from (*5*) with permission]. Horizontal scale bar is 20 ms; vertical scale bar is 20 mv and 1 mv, respectively, for intracellular and nerve records.

---

[1]Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853, USA. [2]Department of Anatomy, Howard University College of Medicine, Washington, DC 20059, USA. [3]Department of Physiology and Neuroscience, New York University Medical Center, New York, NY 10016, USA. [4]Marine Biological Laboratory, Woods Hole, MA 02543, USA. [5]Bodega Marine Laboratory, University of California, Bodega Bay, CA 94923, USA.

*To whom correspondence should be addressed. E-mail: ahb3@cornell.edu

at the same stage showed vocal prepacemaker neurons and the rostral pacemaker–motor neuron circuit in caudal rh8. The caudal pacemaker–motor circuit extended into the rostral spinal cord. The dorsoventral position of the vocal pacemaker circuit was corroborated in transverse series (Fig. 4E); other experiments verified its location relative to nonvocal motor neurons. Because biotin dextran-amine is not transported transneuronally (*6*, *16*), both VMN and a contralateral, nonvocal subgroup could be labeled in the same larva. As observed at early stages (Fig. 3), the caudal part of the VMN was coextensive with occipital-spinal neurons innervating nonvocal muscles [EP in (Fig. 4F)]. In sum, the topography of vocal and nonvocal neurons in larvae with a pacemaker circuit showed that both vocal pacemaker and motor neurons develop in the same rh8-spinal compartment.

We also identified the inferior olive, a precerebellar nucleus, because it is derived from rh8 in tetrapods (*13*) and, thus, provides a critical, non–motor neuron landmark for establishing the position of the developing vocal circuit within the caudal hindbrain. Neurobiotin labeling of the larval cerebellum showed the inferior olive's extent coincident with the vocal prepacemaker and rostral pacemaker–motor circuit (Fig. 4, G and H). Together, the combinatorial mapping of vocal neurons with inferior olive, reticulospinal, and nonvocal motor neurons firmly established the location of the entire vocal pacemaker–motor circuit of larval fish within a compartment that spans hindbrain rh8 and the rostral spinal cord (Fig. 4I).

A taxonomic survey revealed a close alignment of the larval fish pattern with the vocal circuitry of other groups of vocal vertebrates (Fig. 4J). The position of the VMN compares well with that of developing vocal motor neurons that innervate the syrinx in birds (XIIts) (*13*) and larynx in frogs (XMNc) (*17*). Although comparable developmental studies have not been done for reptiles and mammals, their adult phenotypes indicate a similar pattern for laryngeal motor neurons (Amb) (*18*, *19*). Premotor neurons that pattern vocal-respiratory mechanisms in birds (RAm/PAm) (*20*, *21*), amphibians (Ri) (*22*), and mammals (*23*, *24*), including nonhuman primates and humans (Drt/RAb) (*25*), are topographically comparable to the developing prepacemaker-pacemaker nuclei (VPP-VPN) of fish and, like the VPP-VPN (*4*), are targets of midbrain vocal neurons (*24*).

The most parsimonious interpretation of a cladistic analysis (*10*) is that a vocal hindbrain-spinal compartment originated in a common ancestor of the two major groups of living fishes, the Actinopterygii, which includes the teleosts studied here, and the Sarcopterygii, which includes lobe-finned fishes (lungfish and coelacanth) and tetrapods (Fig. 1A). The vocal muscles of fish and tetrapods also share origins from occipital somites (Fig. 1A) (*7*, *26*, *27*), although the innervation of the vocal muscle associated with the sound-generating organ evolved at least three times (Fig. 1A): occipital nerve and swim bladder of fish (Actinopterygii) (*4*) (fig. S1); vagal nerve and larynx of nonavian tetrapods (*17*–*19*, *22*–*25*); and hypoglossal nerve and syrinx of birds (*20*).

We conclude that our results reveal the ancestral origins of neural pattern generators for vocal-acoustic behaviors that mediate social signaling among all the major vertebrate lineages (Fig. 1A). The phylogenetically robust presence of a vocal compartment that shares an organization comparable to cranial motor nuclei predicts its developmental specification by ancestrally conserved, genetic regulatory pathways including a combination of 5′ Hox gene expression in both mesoderm (in this case, occipital somites) and the region spanning the developing hindbrain and spinal cord (*28*). Avian studies emphasize similarities with mammals in forebrain audio-vocal pathways (*1*). We propose that a more ancient neuroectodermal compartment, originating before the radiation of fishes that includes tetrapods, gave rise to the premotor-motor circuitry providing timing signals for a wide range of vertebrate acoustic behaviors (Fig. 1). As Darwin [(*29*) pp. 23, 331] commented regarding the "musical" sounds of male fish during the breeding season: "when the primeval members

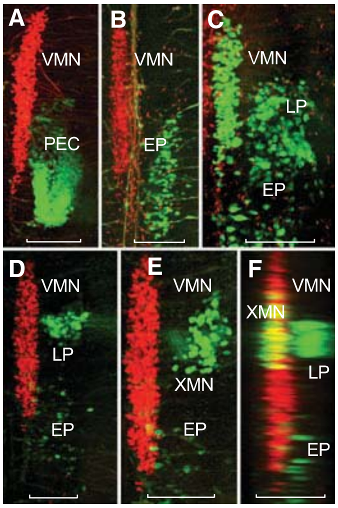


Fig. 3. Segmental organization of vocal and nonvocal motor neurons. Simultaneous labeling of vocal and either nonvocal muscles or nerve roots with different wavelength fluorescent dextran-amines provides visualization of vocal motor neurons (VMNs) and nonvocal motor neurons in midshipman larvae (~14 dpf, *n* = 28). Confocal projections in horizontal (A to E) and sagittal (F) planes. Nonvocal motor neurons are designated by target muscle: EP, epaxialis; LP, levator pectoralis; PEC, pectoral fin. XMN indicates vagal motor neurons innervating wall of gastrointestinal tract. Neurons appearing red and green were labeled with, respectively, fluorescent dextran-amines 546 and 488. Scale bars are 0.1 mm. (**A** and **B**) PEC and EP neurons lateral to caudal vocal motor neurons. (**C**) LP and EP lateral to caudal vocal motor neurons after intracranial label of ipsilateral occipital roots; partial view of contralateral vocal motor neurons filled (far left) by labeling contralateral occipital root. (**D**) LP neurons lateral to vocal motor neurons with sparsely distributed EP neurons labeled during the exposure. (**E**) Higher magnification view just dorsal to (D) showing XMN lateral, rather than directly dorsal, to vocal motor neurons because of tissue flat mounting (see fig. S1). (**F**) Sagittal reconstruction of (D) and (E) showing XMN and LP relative to more medial vocal motor neurons. Yellow is composite overlap (red and green), but no double label.

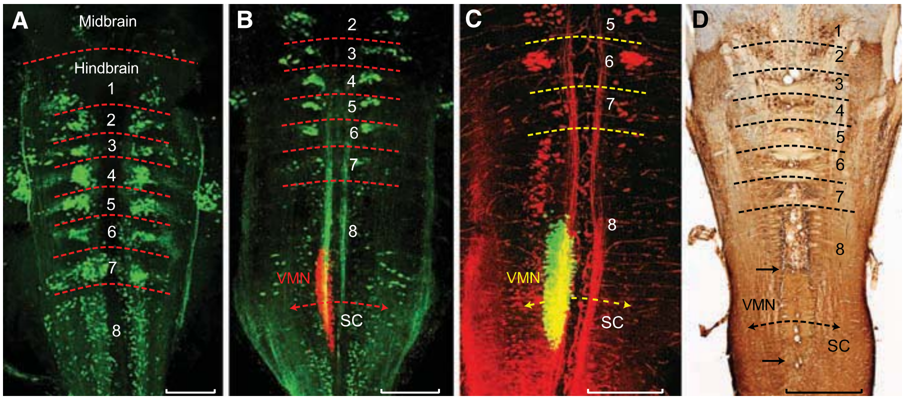


Fig. 2. Segmental organization of hindbrain reticulospinal and vocal neurons. (A to C) Confocal projections in horizontal plane of fluorescently labeled neurons in midshipman larvae [*n* = 21; 8 to 14 days post fertilization (dpf), 7- to 10-mm standard length]. Horizontal hatching approximates midbrain, hindbrain rhombomere (1 to 8), and spinal cord (SC) boundaries. Scale bars are 0.2 mm. (**A**) Clusters of reticulospinal neurons labeled with Alexa biocytin 488 (green) at a stage before vocal muscle reaches swim bladder (~8 dpf). (**B**) Simultaneous visualization of reticulospinal neurons and vocal motor neurons (VMNs) labeled with, respectively, Alexa 488 dextran-amine (green) and Alexa 546 dextran-amine (red) at a stage when vocal muscles attach to the swim bladder (~14 dpf). (**C**) Same stage as (B) highlights rostral subdivision of rh8 reticulospinal column (Alexa 546 dextran-amine, red) and VMN (Alexa 488 dextran-amine, green). Yellow in (B) and (C) is composite overlap (red and green), but no double label. (**D**) Horizontal section showing biocytin-filled, reticulospinal clusters in postlarval midshipman (80 mm, ~200 dpf). Paired black arrows denote VMN's rostral-caudal extent. Scale bar is 0.5 mm.

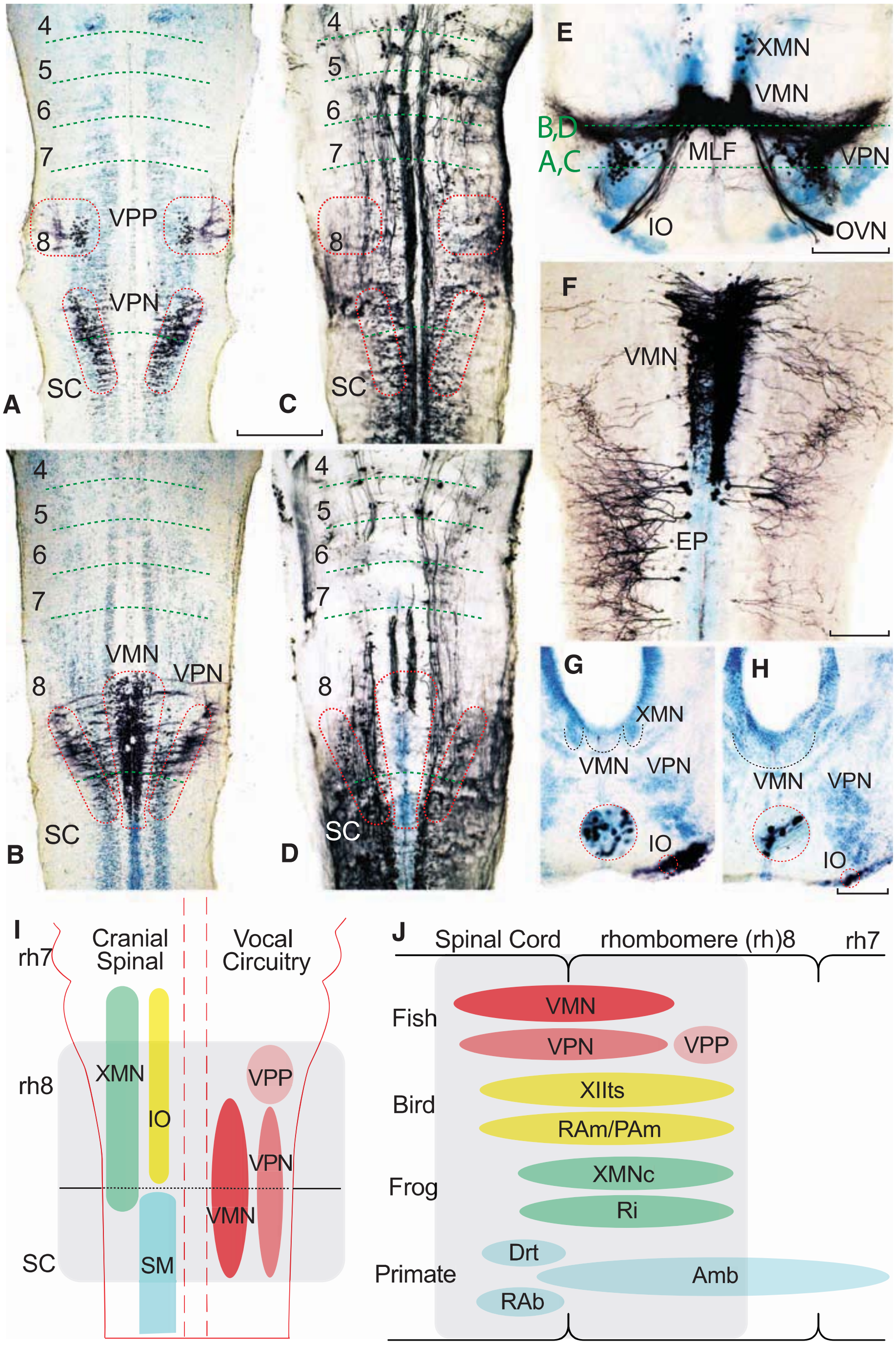


Fig. 4. Segmental organization of vocal pacemaker–motor neuron network along with summary diagrams of larval fish and vocal vertebrates. (A to H) Alignment of either neurobiotin- or dextran-biotin–filled (which appear black) vocal and nonvocal neurons in toadfish (A to F) and midshipman (G and H) larvae (*n* = 25; 1.8 to 2.0 cm, ~30 to 40 dpf). Cresyl violet counterstain illustrates other neuronal groups. Paired horizontal sections at two dorsal-ventral levels, indicated in (E), compare neurobiotin-filled vocal prepacemaker (VPP), pacemaker (VPN), and motor (VMN) neurons (**A** and **B**) relative to neurobiotin-filled reticulospinal neurons (**C** and **D**). Red hatching in (C) and (D) approximates location of the VPP, VPN, and VMN in (A) and (B). Horizontal hatching in (A to D) estimates rhombomere (4 to 8) and spinal cord (SC) boundaries. Scale bar is 0.2 mm for (A to D). (**E**) Transverse section through rostral, neurobiotin-filled vocal pacemaker–motor neurons and exiting occipital vocal tract and nerve root (OVN), including a few labeled vagal motor neurons (XMN) that innervate swim bladder smooth muscle (*15*). (IO, inferior olive; MLF, medial longitudinal fasciculus.) Scale bar is 0.1 mm. (**F**) Horizontal section of VMN (right side) and trunk (epaxialis, EP; left side) motor neurons labeled with 10-kD biotin-dextran-amine. VMN exhibits dendritic spread ipsilaterally and into contralateral VMN. Scale bar is 0.1 mm. (**G** and **H**) Transverse sections showing neurobiotin-labeled inferior olive (IO) coextensive with rostral (G) and middle (H) VMN levels (cresyl violet–stained VMN, outlined with black dotted line, is adjacent to XMN, also outlined, and/or VPN). Red-ringed inserts are higher magnification views of IO. Scale bars are 0.1 and 0.03 mm. (**I**) Schematic in horizontal plane showing relative positions of vocal and nonvocal neurons comprising a rhombomere (rh) 8 to spinal (SC) compartment. (**J**) Comparative summary in sagittal plane showing relative positions of vocal neurons in the rh8–spinal compartment. The developmental map for premotor-motor circuitry in fish (this report), as well as for motor neurons in birds (*13*) and frogs (*17*), is based on representatives of similar age. Positioning of premotor neurons that participate in vocal patterning in birds (*20*, *21*) and frogs (*22*), and motor and premotor neurons in mammals (*19*, *23*, *24*), including primates (*25*), is based on adult phenotypes [but see (*8*) for developmental studies of caudal hindbrain in mammals]. Most laryngeal motor neurons that shape the temporal envelope of mammalian calls originate from caudal nucleus ambiguus (Amb) (*23*). Abbreviations: Drt, dorsal reticular nucleus; PAm, nucleus parambigualis; Ri, inferior reticular formation; RAb, nucleus retroambiguus; RAm, nucleus retroambigualis; VPP-VPN-VMN, vocal prepacemaker–pacemaker–motor neurons; XMNc, caudal XMN; XIIts, tracheosyringeal division of hypoglossal motor nucleus.

of this class [Vertebrata] were strongly excited and their muscles violently contracted, purposeless sounds would almost certainly have been produced; and these, if they proved in any way serviceable, might readily have been modified or intensified by the preservation of properly adapted variations."

References and Notes

1. S. A. White, S. E. Fisher, D. H. Geschwind, C. Scharff, T. E. Holy, *J. Neurosci.* **26**, 10376 (2006).
2. F. Vargha-Khadem, D. G. Gadian, A. Copp, M. Mishkin, *Nat. Rev. Neurosci.* **6**, 131 (2005).
3. J. S. Nelson, *Fishes of the World* (Wiley, Hoboken, NJ, ed. 4, 2006).
4. A. H. Bass, J. R. McKibben, *Prog. Neurobiol.* **69**, 1 (2003).
5. A. H. Bass, R. Baker, *J. Neurobiol.* **21**, 1155 (1990).
6. A. H. Bass, M. A. Marchaterre, R. Baker, *J. Neurosci.* **14**, 4025 (1994).
7. H. C. Tracy, *J. Comp. Neurol.* **111**, 27 (1959).
8. E. Gilland, R. Baker, *Brain Behav. Evol.* **66**, 234 (2005).
9. E. Hanneman, B. Trevarrow, W. K. Metkalfe, C. B. Kimmel, M. Westerfield, *Development* **103**, 49 (1988).
10. Materials and methods are available as supplemental material on *Science* Online.
11. A. H. Bass, R. Baker, *Brain Behav. Evol.* **50** (suppl. 1), 3 (1997).
12. S. Guthrie, *Nat. Rev. Neurosci.* **8**, 859 (2007).
13. F. Cambronero, L. Puelles, *J. Comp. Neurol.* **427**, 522 (2000).
14. R. A. Winterbottom, *Proc. Acad. Nat. Sci. Philadelphia* **125**, 225 (1974).
15. L. E. Morita, T. E. Finger, *J. Comp. Neurol.* **264**, 231 (1987).
16. R. Knapp, M. A. Marchaterre, A. H. Bass, *J. Neurobiol.* **38**, 475 (1999).

17. H. Straka, R. Baker, E. Gilland, *J. Comp. Neurol.* **494**, 228 (2006).
18. M. C. Kennedy, *Brain Res.* **218**, 337 (1981).
19. S. Kitamura, J. Okubo, K. Ogata, A. Sakai, *Exp. Neurol.* **97**, 592 (1987).
20. J. M. Wild, *Ann. N. Y. Acad. Sci.* **1016**, 438 (2004).
21. M. F. Kubke, Y. Yazaki-Sugiyama, R. Mooney, J. M. Wild, *J. Neurophysiol.* **94**, 2379 (2005).
22. E. Zornik, D. B. Kelley, *J. Comp. Neurol.* **501**, 303 (2007).
23. R. Rubsamen, H. Schweizer, *J. Comp. Physiol. [A]* **159**, 689 (1986).
24. U. Jürgens, *Neurosci. Biobehav. Rev.* **26**, 235 (2002).
25. U. Jürgens, L. Ehrenreich, *Brain Res.* **1148**, 90 (2007).
26. D. M. Noden, P. Francis-West, *Dev. Dyn.* **235**, 1194 (2006).
27. R. Huang, Q. Zhi, J.-C. Izpisua-Belmonte, B. Christ, K. Patel, *Anat. Embryol. (Berlin)* **200**, 137 (1999).
28. J. Deschamps, J. van Nes, *Development* **132**, 2931 (2005).
29. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (reprint of 1871 edition by John Murray, London, Princeton Univ. Press, Princeton, NJ, 1981).
30. B. A. Moore, A. P. Russell, A. M. Bauer, *J. Morphol.* **210**, 227 (1991).
31. Thanks to D. Chiu, D. Brown, I. Fein, A. Mensinger, L. Remage-Healey for logistical and technical support; K. Zamudio for help with the cladogram; M. Nelson for drawings; M. Marchaterre for movies; A. Rice for movie editing; E. Adkins-Regan, G. Budney, U. Jürgens, M. Marchaterre for vocal recordings; J. Fetcho, H. Greene, L. Z. Holland, C. Watson, and especially H. Baker, R. Hoy, and J. M Wild for comments on the manuscript. Research support from the NSF and NIH.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5887/417/DC1
Materials and Methods
Figs. S1 and S2
References
Movies S1 to S3

11 March 2008; accepted 4 June 2008
10.1126/science.1157632

Orbitofrontal Dysfunction in Patients with Obsessive-Compulsive Disorder and Their Unaffected Relatives

Samuel R. Chamberlain,[1,2,3]* Lara Menzies,[1,2,3] Adam Hampshire,[4] John Suckling,[1,2] Naomi A. Fineberg,[1,3] Natalia del Campo,[1,2] Mike Aitken,[2,5] Kevin Craig,[1,2,3] Adrian M. Owen,[4] Edward T. Bullmore,[1,2,6] Trevor W. Robbins,[2,5] Barbara J. Sahakian[1,2]

Obsessive-compulsive disorder (OCD) is characterized by repetitive thoughts and behaviors associated with underlying dysregulation of frontostriatal circuitry. Central to neurobiological models of OCD is the orbitofrontal cortex, a neural region that facilitates behavioral flexibility after negative feedback (reversal learning). We identified abnormally reduced activation of several cortical regions, including the lateral orbitofrontal cortex, during reversal learning in OCD patients and their clinically unaffected close relatives, supporting the existence of an underlying previously undiscovered endophenotype for this disorder.

Obsessive-compulsive disorder (OCD) is a debilitating neuropsychiatric condition characterized by recurrent intrusive thoughts (obsessions) and/or repetitive rituals (compulsions), often performed according to rigid rules (*1, 2*).

OCD is frequently familial, and first-degree relatives of patients exhibit up to an eightfold increased risk of also developing clinically meaningful OC symptoms (*3*). Nonetheless, attempts to delineate contributory genes have met with limited success. It is probable that top-level symptoms are too distal from the underlying pathogenesis of the disorder to provide sufficient power to detect underlying genetic diatheses. Consequently, there is an ongoing search for objective brain-based measurable traits, or endophenotypes, that decompose top-level phenotypes into meaningful markers more proximally related to the etiology (*4*, *5*). Such markers should be present in unaffected first-degree relatives of patients, even in the absence of clinically meaningful symptoms (*4*).

OCD is associated with abnormal function in corticostriatal circuitry mediating inhibitory control and flexible responding (*6*, *7*). The orbitofrontal cortex (OFC) is central to our understanding of OCD (*8*) because structural and functional alterations of this region are the most frequently reported neuroimaging findings in patient studies (*6*, *7*, *9*). The OFC subserves reversal learning, a cognitive function important in day-to-day life whereby behavior is flexibly altered after negative feedback (*10*). Reversal learning is dependent on the serotoninergic system (*11*, *12*) and is impaired by lesions to the OFC (but not the dorsolateral prefrontal cortex) across species (*13*, *14*). Reduced activation of the OFC has previously been reported in patients with OCD during reversal learning (*15*), but many patients in that study were also depressed. This was a potential confounder because depression itself is associated with OFC dysfunction and increased feedback sensitivity (*16*). In addition, relatives were not assessed, and thus the state-versus-trait nature of the OFC dysfunction could not be discerned.

In the search for neurocognitive endophenotypes in OCD, we measured brain activation in comorbidity-free patients and their unaffected first-degree relatives. We used a functional magnetic resonance imaging task capable of fractionating different components of behavioral flexibility, including reversal of responses, a plausible psychological deficit in OCD (*17*). The sample comprised 14 patients, 12 unaffected never-treated first-degree relatives of these patients, and 15 matched controls without a family history of the disorder. The three groups were matched in terms of age, handedness, and intelligence quotient (table S1).

On each trial, volunteers observed two pictures presented on screen, each of which comprised a face and house superimposed. The aim of the task was for volunteers to work out through trial and error which object (which face or house) was correct (fig. S1). If volunteers believed that the left-hand stimulus contained the correct object, they pressed a left button, and vice versa. After every second response, feedback was given on screen ("CORRECT" or "INCORRECT") to indicate whether the chosen object was indeed correct. Once a criterion of six consecutive correct responses was reached, either the correct object was changed or a new stimulus set was presented; the volunteer was then required to learn the new correct object (*17*, *18*).

We first examined the volunteers' brain activation when they were working out solutions, minus their brain activation when the solutions were known. This provided an overview of the neural circuitry generally involved in undertaking the task. We then examined their brain activations corresponding to reversal learning and set-shifting, cognitive functions dependent on the OFC and ventrolateral prefrontal cortices, respectively (*10*, *17*). We hypothesized that patients with OCD and their unaffected relatives would show reduced activation of the OFC in the reversal contrast (*6*) as compared with controls. The groups did not differ significantly in number of errors on the task, because all participants were pretrained to minimize the likelihood of behavioral confounders. However, relatives exhibited slower response times when working out solutions than the patients and controls (fig. S2).

Regions activated when working out solutions and during reversal learning, across all subjects, are indicated in fig. S4, A and B. When working out solutions, OCD patients and their unaffected relatives showed under-activation in regions including, bilaterally, the lateral OFC [Brodmann areas (BA) 10, 11, and 47], lateral prefrontal cortex (PFC) (BA 45 and 46), and left

---

[1] Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Box 189, Cambridge CB2 0QQ, UK. [2] Behavioural and Clinical Neuroscience Institute (BCNI), University of Cambridge, Cambridge CB2 3EB, UK. [3] National OCD Treatment Service, Queen Elizabeth II Hospital, Welwyn Garden City AL7 4HQ, UK. [4] Medical Research Council (MRC) Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK. [5] Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK. [6] Clinical Unit Cambridge, GlaxoSmithKline, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.

*To whom correspondence should be addressed. E-mail: srchamb@gmail.com

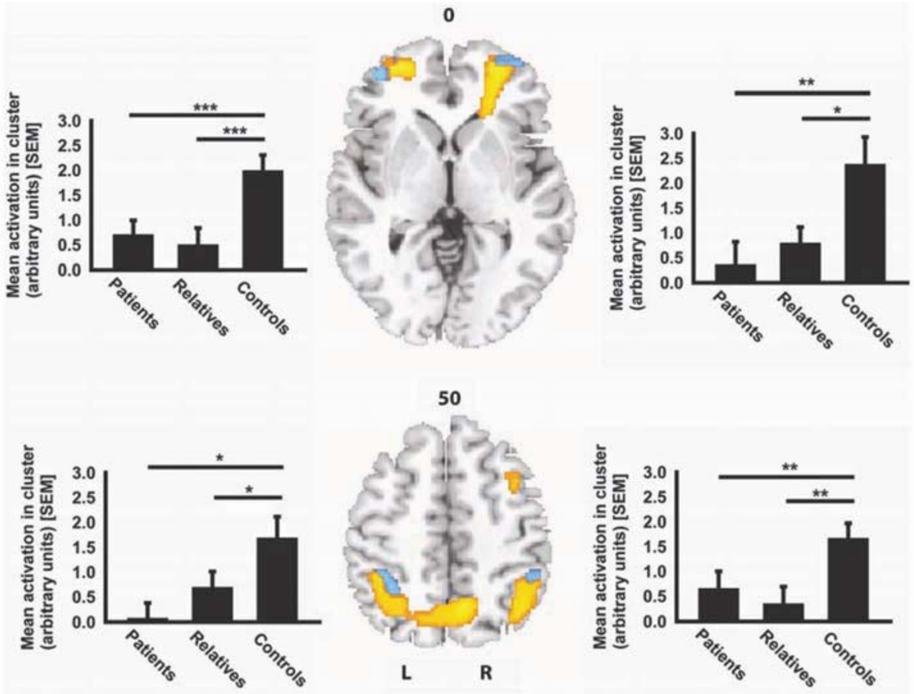


Fig. 1. Patients with OCD and their unaffected relatives showed underactivation during reversal learning bilaterally in the lateral OFC, lateral PFC, and parietal cortices. The images are of representative brain slices showing regions activated during reversal learning across all subjects (yellow areas; false discovery rate–corrected, $P < 0.05$) and regions in which there was a significant effect of group (blue areas; corrected to less than one false-positive cluster across the whole map) (*19*). Peripheral graphs indicate mean group activations for each of the four identified clusters where there was a significant effect of group. (**Top left**) Cluster 1, left lateral OFC (BA 10, 11, and 47) and left lateral PFC (BA 46). (**Top right**) Cluster 2, right lateral OFC (BA 10, 11, and 47) and right lateral PFC (BA 46). (**Bottom left**) Cluster 3, left parietal lobe (BA 40). (**Bottom right**) Cluster 4, right parietal lobe (BA 40). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ are the significant differences in brain activation between groups, using non-parametric permutation.

parietal cortex (BA 40) as compared with controls (*18*). During reversal learning, OCD patients and relatives showed significant underactivation bilaterally in the lateral OFC (BA 10, 11, and 47), lateral PFC (BA 46), and parietal cortex (BA 40) as compared with controls (Fig. 1). Brain activation during extradimensional shifting was not significant at the whole-study level and thus was not entered into between-group analysis (*18*).

This study found reduced lateral OFC, lateral PFC, and parietal responsiveness during reversal learning not only in patients with OCD but also in their unaffected never-treated relatives. These findings emphasize the centrality of these regions not only in day-to-day flexibility but also in the genesis of pathologic habits. Reversal-learning–related hypofunction appears to be a vulnerability marker (or candidate endophenotype) for OCD that exists in people at increased genetic risk, even in the absence of chronic treatment or symptom confounders. Such brain-based markers may thus be of considerable utility in the search for underlying genetic diatheses.

References and Notes

1. DSM-IV, *Diagnostic and Statistical Manual of Mental Disorders IV (DSM-IV)* (American Psychiatric Association, Arlington, VA, 1994).
2. E. Hollander *et al.*, *J. Clin. Psychiatry* **57** (suppl. 8), 3 (1996).
3. D. L. Pauls, J. P. Alsobrook 2nd, W. Goodman, S. Rasmussen, J. F. Leckman, *Am. J. Psychiatry* **152**, 76 (1995).
4. I. I. Gottesman, T. D. Gould, *Am. J. Psychiatry* **160**, 636 (2003).
5. A. Meyer-Lindenberg, D. R. Weinberger, *Nat. Rev. Neurosci.* **7**, 818 (2006).
6. S. R. Chamberlain, A. D. Blackwell, N. Fineberg, T. W. Robbins, B. J. Sahakian, *Neurosci. Biobehav. Rev.* **29**, 399 (2005).
7. L. Menzies *et al.*, *Neurosci. Biobehav. Rev.* **32**, 525 (2008).
8. S. Saxena, R. G. Bota, A. L. Brody, *Semin. Clin. Neuropsychiatry* **6**, 82 (2001).
9. L. Menzies *et al.*, *Brain* **130**, 3223 (2007).
10. R. Dias, T. W. Robbins, A. C. Roberts, *Nature* **380**, 69 (1996).
11. H. F. Clarke *et al.*, *J. Neurosci.* **25**, 532 (2005).
12. S. R. Chamberlain *et al.*, *Science* **311**, 861 (2006).
13. J. Hornak *et al.*, *J. Cogn. Neurosci.* **16**, 463 (2004).
14. V. Boulougouris, J. W. Dalley, T. W. Robbins, *Behav. Brain Res.* **179**, 219 (2007).
15. P. L. Remijnse *et al.*, *Arch. Gen. Psychiatry* **63**, 1225 (2006).
16. S. R. Chamberlain, B. J. Sahakian, *Curr. Psychiatry Rep.* **8**, 458 (2006).
17. A. Hampshire, A. M. Owen, *Cereb. Cortex* **16**, 1679 (2006).
18. Materials and methods are available as supporting material on *Science* Online.
19. J. Suckling, E. T. Bullmore, *Hum. Brain Mapp.* **22**, 193 (2004).
20. Supported in part by a Wellcome Trust Programme Grant (076274/Z/04/Z) to T.W.R., B.J.S., B. J. Everitt, and A. C. Roberts. The BCNI is supported by a joint award from the Medical Research Council and Wellcome Trust (G001354). S.R.C. was supported by a priority studentship from the Medical Research Council. L.M. was supported by the Harnett Fund, University of Cambridge. E.T.B. was supported by a Distinguished Investigator Award from the National Alliance for Research in Schizophrenia and Affective Disorders. A.M.O. and A.H. were supported by MRC (U1055.01.002.00001.01). The authors thank the study participants and radiographers at the Wolfson Brain Imaging Centre, Addenbrooke's Hospital, Cambridge, UK. Software development was supported by a Human Brain Project grant from the National Institute of Mental Health and the National Institute of Biomedical Imaging and Bioengineering.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5887/421/DC1
Materials and Methods
Figs. S1 to S7
Table S1
References

20 December 2007; accepted 17 June 2008
10.1126/science.1154433

New Products Focus: Liquid Handling

Fast Sample Homogenization

The Omni Prep homogenizing station streamlines the sample preparation process and removes tedious) time-consuming manual steps. The six-station system offers high throughput and eliminates cross-contamination. It is completely flexible in terms of tube size) sample type) and volume. It includes timer control) variable speed) and variable vertical oscillation to minimize processing time and handle almost any sample type. Plastic probes) strong enough to be used from the softest sample to difficult-to-homogenize tissues) can be attached to the system along with the more traditional stainless steel tips. The plastic probes are designed for single use to eliminate sample contamination) but can be cleaned and reused if desired.

Camlab
For information +44-01954-233100
www.camlab.co.uk

Concentrators for Protein Prep

The Vivaspin sample concentrators for protein sample preparation offer fast and convenient concentration of biological samples with minimal membrane blocking) recoveries of greater than 95 percent) and no risk of drying out the sample. Vivaspin provides fast) nondenaturing concentration of up to 30 times for volumes from 100 µl to 20 ml. The instrument concentrates samples by membrane filtration to a single tube) with the molecular weight cutoff selected by the user. The sample is placed in the upper compartment of the tube) and centrifugation forces solvent through a semipermeable membrane into the lower compartment. This leaves a concentrated sample in the upper chamber.

GE Healthcare
For information 732-457-8149
www.gelifesciences.com/sampleprep

Label-Free Assay Accessory

A new automated liquid-handling accessory (LHA) integrates with the Corning Epic System for high throughput label-free assays. The Epic LHA includes a 384-well pipettor) a microplate stacker that can handle up to 20 lidded 384-well microplates and an automated tip wash station. The liquid-handling system is designed to maintain thermal stability to eliminate artifacts from temperature fluctuations. The Epic System is a high throughput label-free drug discovery system that features patented optical biosensor technology to perform biochemical and cell-based assays. The system is sensitive enough to measure the cell response from endogenous receptor expression on primary or nonengineered cell lines) along with protein/drug direct binding events in biochemical assays.

Corning
For information 978-442-2275
www.corning.com/lifesciences

Safe Vacuum System

The Vacusafe system was developed for safe) easy) and rapid handling of hazardous liquids. The new generation Vacusafe is designed for aspiration) disposal) or filtration of biological and chemical liquids in a clean and professional way. All required functionality is integrated in a compact design that can be used at the push of a button. For aspiration and disposal applications) the Vacusafe can replace noisy) complicated “home made” systems or systems connected directly to the laboratory vacuum system that risk contamination of the entire vacuum system.

Integra Biosciences
For information +41-81-286-9530
www.vacusafe.info

Multilabel Plate Reader

The Victor X Multilabel Plate Reader offers the flexibility to support new applications beyond primary screening) including quality control and therapeutic research. It includes an enhanced software package that features an improved “Start Wizard” for easier creation of new protocols) application-based example protocols for off-the-shelf use) and improved kinetics support for monitoring cellular and enzyme assays in real time. The Victor X units are also designed to be easy to upgrade with new accessories.

PerkinElmer
For information 781-663-5701
www.perkinelmer.com

50-µl Disposable Tip

A 50-µl size is a new addition to a line of conductive) disposable tips that include sizes of 10-µl) 200-µl) and 1)000-µl. The tips fit Tecan’s automated liquid-handling workstations) enhancing performance and minimizing cross-contamination in a broad range of genomics) forensics) and enzyme-linked immunosorbent assay (ELISA) applications. The 50-µl tips are optimal for polymerase chain reaction set-up) DNA quantification and normalization) and small-volume applications such as ELISAs.

Tecan
For information +41-44-922-81-11
www.tecan.com

Nanoliter Dispensing Tips

A new line of PocketTips enables nanoliter-volume dispensing on laboratory automated systems. An internally molded capillary pocket allows transfer of 50-nl to 500-nl molecular compounds suspended in dimethyl sulfoxide (DMSO) directly to assay plates. This in-tip mixing saves time) reagents) and labware as well as improving data quality. PocketTips are made from high-quality polypropylene that eliminates adherence and maximizes the recovery of fluids. Through aspiration) the capillary pocket fills with the specified amount of compound solution and facilitates soft in-tip mixing of sample and assay material. By removing intermediate dilution steps) this innovative pipetting method makes it possible to perform assays more quickly) conserve valuable compounds) and prevent compound precipitation for more accurate and precise results. The tips are suitable for all secondary compound screening applications) such as selectivity assays) potency determinations) cell- and enzyme-based assays) and polymerase chain reaction primer dilutions.

Thermo Fisher
For information 508-742-5254
www.thermofisher.com

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.